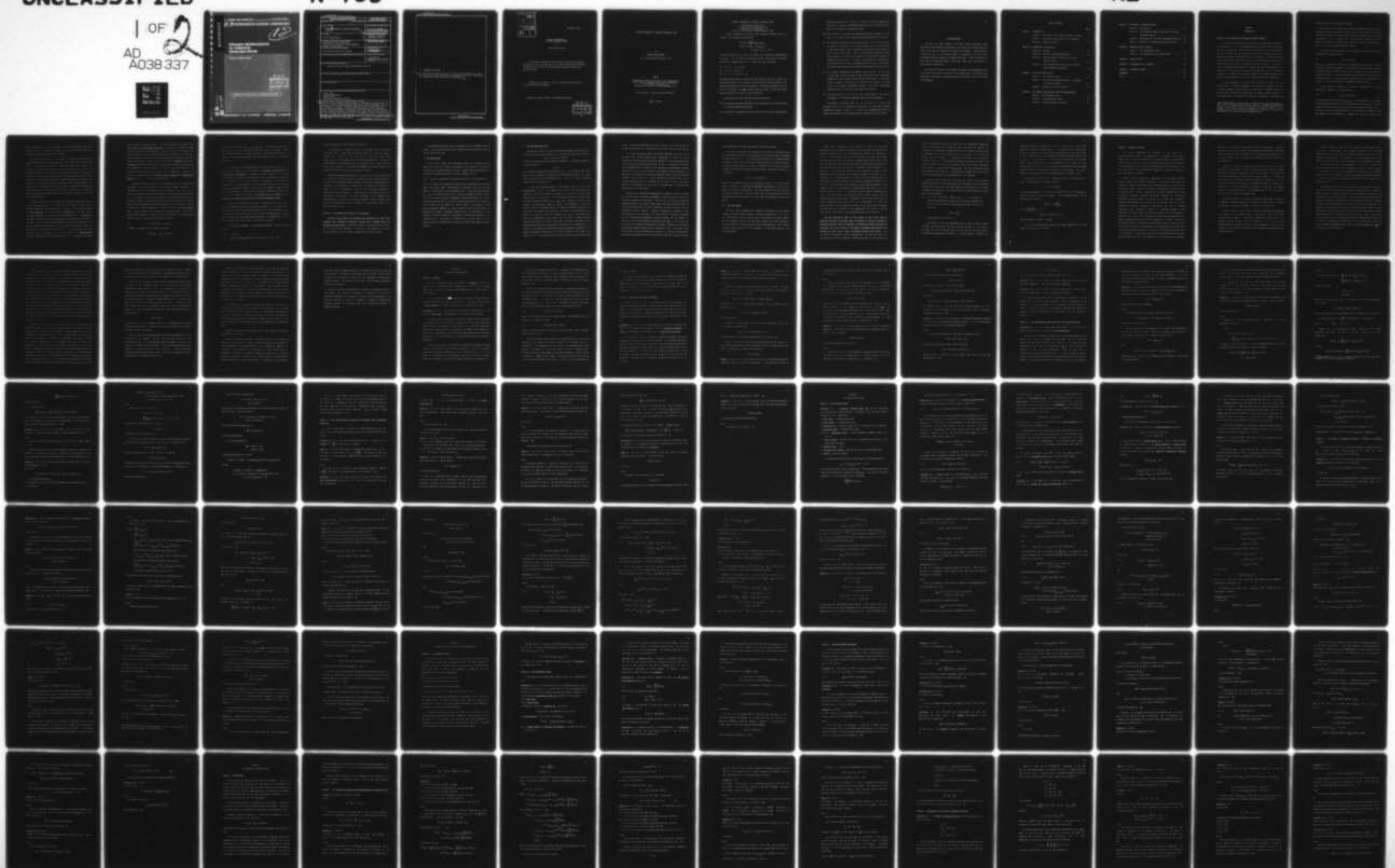
AD-A038 337 ILLINOIS UNIV AT URBANA-CHAMPAIGN COORDINATED SCIENCE LAB F/G 12/2
DYNAMIC PROGRAMMING IN COMPLETE SEPARABLE SPACES.(U)
JAN 77 S E SHREVE DAAB07-72-C-0259
UNCLASSIFIED R-755 NL

1 OF 2
AD
A038337



337

REPORT R-755 JANUARY, 1977

AD A038337

CSL COORDINATED SCIE

DYNAMIC PROGRAMMI
IN COMPLETE
SEPARABLE SPACES

STEVEN EUGENE SHREVE

UNIVERSITY OF ILLINOIS –

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle)	DYNAMIC PROGRAMMING IN COMPLETE SEPARABLE SPACES
5. TYPE OF REPORT & PERIOD COVERED	Technical Report
6. PERFORMING ORG. REPORT NUMBER	R-755, UILU-ENG-77-2202
7. AUTHOR(s)	Steven Eugene Shreve
8. CONTRACT OR GRANT NUMBER(s)	DAAB-07-72-C-0259; NSF-ENG-74-19332
9. PERFORMING ORGANIZATION NAME AND ADDRESS	Coordinated Science Laboratory University of Illinois at Urbana-Champaign Urbana, Illinois 61801
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	153 p.
11. CONTROLLING OFFICE NAME AND ADDRESS	Joint Services Electronics Program
12. REPORT DATE	January, 1977
13. NUMBER OF PAGES	143
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	
15. SECURITY CLASS. (of this report)	UNCLASSIFIED
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Borel Space
Dynamic Programming
Analytic Set
Universally Measurable Policies

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The discrete time stochastic decision model is a mathematical abstraction of the situation in which a system progresses from state to state incurring a cost at each transition. The cost could be assigned to reflect the preference one has for one state over another or could be the genuine cost of say, operating a business, during the period of transition. A decision maker has some influence over the stochastic manner of the transition but cannot choose deterministically the state into which the system will move. He wishes, of course, to exercise his influence to minimize the total expected cost of all transitions. Thus he must

next page

DD FORM 1473, 1 JAN 73 — EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 097 700

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (continued)

cont

not only take into account the cost of the present transition, but rather must balance his desire to minimize this against his desire to avoid moving to a state where a high future cost is unavoidable.

UILU-ENG 77-2202

DYNAMIC PROGRAMMING IN COMPLETE SEPARABLE SPACES

by

Steven Eugene Shreve

This work was supported in part by the Joint Services Electronics Program (U.S. Army, U.S. Navy and U.S. Air Force) under Contract DAAB-07-72-C-0259 and in part by the National Science Foundation under Grant NSF ENG 74-19332.

DYNAMIC PROGRAMMING IN COMPLETE SEPARABLE SPACES

BY

STEVEN EUGENE SHREVE

A.B., West Virginia University, 1972

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mathematics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1977

Thesis Advisor: Professor Dimitri Bertsekas

Urbana, Illinois

DYNAMIC PROGRAMMING IN COMPLETE SEPARABLE SPACES

Steven Eugene Shreve, Ph.D
Coordinated Science Laboratory and
Department of Mathematics
University of Illinois at Urbana-Champaign, 1977

A general framework for discrete time stochastic optimal control is proposed. The sequential decision model

$$\text{minimize: } E\{\sum_{k=0}^{N-1} \alpha^k g(x_k,u_k,w_k)\},$$
$$\text{subject to: } x_{k+1} = f(x_k,u_k,w_k),$$
$$u_k \varepsilon U(x_k),\ k=0,1,\dots,N-1,$$

is treated, where w_k is a random disturbance with distribution parameterized by (x_k,u_k). If N is finite, we define and consider models which are summable below (F^+) or summable above (F^-). If $N=+\infty$, we treat the cases:

(P) $0 < \alpha \leq 1$, $0 \leq g \leq +\infty$,

(N) $0 < \alpha \leq 1$, $-\infty \leq g \leq 0$,

(D) $0 < \alpha < 1$, $-b \leq g \leq b < +\infty$.

The minimization problem is shown to be well posed if the state, control and disturbance spaces are Borel spaces, the other data are Borel measurable, and universally measurable policies are admitted. Universally measurable policies which are ϵ-optimal for <u>every</u> initial state are shown to exist and simple characterizations of optimal policies are provided.

In particular, we have under the indicated conditions:

(F^+) An ϵ-optimal nonrandomized Markov policy exists and can be constructed by the dynamic programming algorithm.

(F^-) An ϵ-optimal (randomized) Markov policy and an ϵ-optimal nonrandomized

semi-Markov policy exist. If $\{\epsilon_n\}$ is a sequence of positive numbers with $\epsilon_n \downarrow 0$, then a sequence of nonrandomized Markov policies exhibiting "$\{\epsilon_n\}$ dominated convergence to optimality" exists.

($F^{\pm}$) If the infimum in the dynamic programming algorithm is achieved for each state at each stage, then an optimal nonrandomized Markov policy exists.

(P) An ϵ-optimal nonrandomized Markov policy exists. If $\alpha < 1$, this policy can be taken to be stationary. If for each initial state, a policy optimal at that state exists, then a nonrandomized stationary policy optimal at every initial state exists. Such a nonrandomized stationary optimal policy exists if and only if the infimum in the optimality equation is achieved for every state. Continuity and compactness conditions are given under which the dynamic programming algorithm yields, in the limit, the optimal cost function and an optimal nonrandomized stationary policy.

(N) An ϵ-optimal nonrandomized semi-Markov policy exists. If for each initial state, a policy optimal at that state exists, then a semi-Markov (randomized) policy optimal at every initial state exists. A stationary policy is optimal if and only if its associated cost function is a fixed point of the dynamic programming operator. The dynamic programming algorithm yields, in the limit, the optimal cost function.

(D) All results given for (P) and (N) hold. Sharp bounds on the rate of convergence of the dynamic programming algorithm are established.

The method of analysis under (P), (N) and (D) is to convert the stochastic model to an equivalent deterministic one and apply standard deterministic results. Finally, it is shown how nonstationary models and models with imperfect state information can be reduced to the one treated.

ACKNOWLEDGEMENTS

The author is deeply indebted to Professor Dimitri Bertsekas, whose knowledge of what should be investigated, what ought to be possible and what others have done was invaluable in focusing the research. Several specific theorems as well as the overall manner of presentation of this thesis incorporate ideas and insights of Professor Bertsekas. A debt of gratitude is also owed to Professors Michael Pursley and Joseph Doob, who pointed out several helpful references.

The author also wishes to acknowledge the support and encouragement given by his wife Dot during the period of his graduate study. Her careful hand lettering of the symbols in this thesis made it possible to meet the deadline for submission.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Section 1. The discrete time stochastic decision problem.

The discrete time stochastic decision model is a mathematical abstraction of the situation in which a system progresses from state to state incurring a cost at each transition. The cost could be assigned to reflect the preference one has for one state over another or could be the genuine cost of say, operating a business, during the period of transition. A decision maker has some influence over the stochastic manner of the transition but cannot choose deterministically the state into which the system will move. He wishes, of course, to exercise his influence to minimize the total expected cost of all transitions.[1] Thus he must not only take into account the cost of the present transition, but rather must balance his desire to minimize this against his desire to avoid moving to a state where a high future cost is unavoidable.

A classical example of this situation, in which we treat profit as negative cost, is portfolio management. An investor must balance his desire to achieve immediate return, possibly in the form of dividends, against a desire to avoid investments in areas where low long-run yield is probable. If the total value of the portfolio is taken as the state of the system, the

---

[1]Many authors speak of maximization of reward rather than minimization of cost. We follow the practice of control theorists. Although this discrepancy involves only a change in sign, one must take care to avoid confusion. For example, if the cost function is nonnegative, we have a "positive dynamic programming model," which is treated extensively in Strauch [35], where it is referred to as a "negative dynamic programming model."

stochastic nature of the problem is apparent.

Other examples can be drawn from inventory management, reservoir control, sequential analysis (hypothesis testing) and, by discretizing a continuous problem, from control of a large variety of physical systems subject to random disturbances. For an extensive set of discrete time stochastic decision models, see Bellman [2], Bertsekas [3], Dynkin and Juskevic [12], Howard [17], Wald [37], and the references contained therein.

If a system is deterministic, then the transition from state to state can be described by a system equation

$$(1.1) \qquad x_{k+1} = f(x_k, u_k),$$

where x_k, x_{k+1} represent a state and its succeeding state and will be assumed to belong to some state space S; u_k represents a control variable chosen by the decision maker in some constraint set $U(x_k)$ which is in turn a subset of some control space C. The cost incurred by such a transition can be given by a function $g(x_k, u_k)$. Actually the cost could depend on x_{k+1} as well, but in view of (1.1) this can be reduced to dependence on only x_k and u_k.

If the system is stochastic we include a disturbance w_k in this description. Equation (1.1) is replaced by

$$(1.2) \qquad x_{k+1} = f(x_k, u_k, w_k),$$

and the cost per stage becomes $g_k'(x_k, u_k, w_k)$. The disturbance w_k is a member of some probability space $(W, \mathfrak{F})$ and has distribution $p(dw_k | x_k, u_k)$. Thus the control variable u_k exercises influence over the transition from x_k to x_{k+1} in two places, once in the system equation (1.2) and again as a parameter in the distribution of the disturbance w_k. Likewise the control u_k influences the

cost at two points. This is a redundancy in the system equation model given above and will be eliminated in Chapter 3 when the transition kernel and reduced one-stage cost function are introduced.

The system equation model is more common in engineering literature and generally more convenient in applications, so we are taking it as our starting point. The transition kernel and reduced one-stage cost function are technical devices which eliminate the disturbance space $(W,\mathcal{F})$ from consideration and make the model more suitable for analysis. We take pains initially to point out how properties of the original system carry over into properties of the transition kernel and reduced one-stage cost function (see Definition 3.2 and following remarks, Theorem 3.5 and following remarks). In Chapter 6 we do not repeat this process but rather introduce the nonstationary and imperfect state information models directly in terms of the transition kernel and reduced one-stage cost function, leaving the reader to infer the system equation models.

To place our model in the literature on stochastic decision theory, we review some terminology. In our model the distribution of state x_{k+1} is entirely determined by the distribution of (x_k,u_k). Such a decision process is called <u>Markovian</u>. The cost structure is <u>additive</u>, i.e. the total cost of the system operation is the sum of all the one-stage costs. The <u>horizon</u> is the number of stages for which the system operates and can be either finite or infinite. We will allow the decision maker to utilize full knowledge of the system structure (the functions f and g and the disturbance distribution p) and, when he is choosing control u_k, he will also know the past states and controls $(x_0,u_0,\ldots,u_{k-1},x_k)$. This is the case of a <u>nonanticipative</u> information structure; the decision maker does not know the next state until

he has chosen the current control. The decision maker does, however, have <u>total</u> <u>recall</u> and <u>perfect</u> <u>state</u> <u>information</u>, i.e. once he has observed a state or chosen a control, this knowledge is not lost to him, and he observes the states accurately. We shall show that despite this abundance of information, the best control possible can be achieved by taking into account only the most recent and perhaps the initial state. This model is <u>stationary</u> (also called <u>homogeneous</u>) because the functions f and g and the disturbance distribution p are independent of the time index k. We hasten to add that these last two conditions on our model really involve no loss of generality. Chapter 6 is devoted to showing that both the <u>imperfect</u> <u>state</u> <u>information</u> and <u>nonstationary</u> <u>models</u> can be reduced to the one considered here.

Stochastic sequential control is distinguished from its deterministic counterpart by the concern with when information becomes available. In deterministic control, a sequence of control variables $(u_0,\ldots,u_{N-1})$ can be specified before-hand and the resulting states of the system are determined by (1.1). In contrast, if the control variables are specified before-hand for a stochastic system, the decision maker may realize in the course of the system evolution that unexpected states have appeared and the specified control variables are no longer appropriate. Thus we are led to consider <u>policies</u> $\pi=(\mu_0,\ldots,\mu_{N-1})$, where μ_k is a function from history to control. If x_0 is the initial state, $u_0=\mu_0(x_0)$ is taken to be the first control. If the states and controls $(x_0,u_0,\ldots,u_{k-1},x_k)$ have occurred, the control

$$u_k = \mu_k(x_0,u_0,\ldots,u_{k-1},x_k) \tag{1.3}$$

is chosen. We require that the control constraint

$$\mu_k(x_0,u_0,\ldots,u_{k-1},x_k) \in U(x_k)$$

be satisfied for every $(x_0, u_0, \ldots, u_{k-1}, x_k)$ and k. In this way the decision maker utilizes the full information available to him at each stage. Rather than choosing a sequence of control variables, the decision maker attempts to choose a policy which minimizes the total expected cost of the system operation.

The analysis of the stochastic decision model outlined above can be pretty much divided into two categories, <u>structural considerations</u> and <u>measurability considerations</u>. Structural analysis consists of all those results which can be gotten if measurability of all functions and sets arising in the problem is of no real concern; for example, if the model is deterministic or, more generally, if the disturbance space W is countable. Measurability analysis consists of showing that the structural results remain valid even when one is forced to place nontrivial measurability restrictions on the set of admissible policies. <u>The present work is primarily one of measurability analysis relying heavily on existing structural results</u>.

One can best illustrate this dichotomy of analysis by the finite horizon dynamic programming algorithm considered by Bellman [2]. The algorithm is based on the intuitively appealing principle that if a policy $\pi=(\mu_0,\ldots,\mu_{N-1})$ is optimal for the N-stage model, then the policies $\pi^k=(\mu_{N-k},\mu_{N-k+1},\ldots,\mu_{N-1})$ must be optimal for the k-stage truncated problem. Put another way, a policy can be optimal only if every "tail" of the policy is optimal.

This observation suggests a computational procedure. Define for every state x,

(1.4) $\qquad J_0(x) = 0,$

(1.5) $\qquad J_{k+1}(x) = \inf_{u\in U(x)} E\{g(x,u,w) + J_k[f(x,u,w)]\}, \; k=0,\ldots,N-1,$

where the expectation is with respect to $p(dw|x,u)$.

It is reasonable to expect that $J_k(x)$ is the optimal cost of operating the system over k stages when the initial state is x, and that if $\mu_k(x)$ achieves the infimum in (1.5) for every x and k=0,...,N-1, then $\pi=(\mu_0,\ldots,\mu_{N-1})$ is an optimal policy for every initial state x. If there are no measurability considerations, this is indeed the case. Notice how the right side of equation (1.5) balances the immediate cost g(x,u,w) against the optimal "cost-to-go" $J_k[f(x,u,w)]$ in choosing a policy.

The dynamic programming algorithm of (1.4) and (1.5) is the simplest and most widely used structural result of stochastic decision theory, and much of our effort will be directed toward proving its validity in a measure theoretic framework. The difficulty, of course, lies in showing the expression in braces in (1.5) is measurable. Thus we must establish measurability properties for the functions J_k. Related to this is the need to balance the measurability restrictions on policies (necessary so the expected cost corresponding to a policy can be defined) against a desire to admit enough policies to consideration so as to be able to find one which selects at or near the infimum in (1.5).

Section 2. The present work related to the literature.

<u>The goal of this thesis is to establish the suitability of a Borel space framework with universally measurable policies for a general theory of stochastic decision models</u>. We show that almost every known structural result can be proved in this framework. In particular, the existence of policies which are optimal or nearly optimal for <u>every</u> initial state is shown.

A great many authors have dealt with measurability in stochastic decision theory. This section describes three approaches taken and how their aims and results relate to our own.

I. The General Model

If the state, control and disturbance spaces are arbitrary measure spaces, very little can be done. One attempt in this direction is the work of Striebel [36] involving p-essential infima. Geared toward giving meaning to the dynamic programming algorithm, this work replaces (1.5) by

(1.6) $$J_{k+1}(x) = p_k\text{-essential infimum}_{\mu}\, E\{g'(x,\mu(x),w) + J_k[f(x,\mu(x),w)]\},$$

$k=0,\ldots,N-1$, where the p-essential infimum is over all measurable μ from state space S to control space C satisfying any constraints which may have been imposed. The functions J_k are measurable, and if the probability measures $p_0,\ldots,p_{N-1}$ are properly chosen and the so-called countable ϵ-lattice property holds, this modified dynamic programming algorithm generates the optimal cost-to-go functions and can be used to obtain policies which are optimal or nearly optimal for p_{N-1}-almost all initial states. The selection of the proper probability measures $p_0,\ldots,p_{N-1}$, however, is at least as difficult as executing the dynamic programming algorithm and the verification of the countable ϵ-lattice property is equivalent to proving the existence of an ϵ-optimal policy. In contrast to Striebel's work, we will impose a Borel space structure on the model which enables us to obtain significantly stronger results.

II. The Semicontinuous Models

Considerable attention has been directed toward models in which the state and control spaces are Borel spaces or even R^n, and the reduced cost function

$$g(x,u) = \int g'(x,u,w)p(dw|x,u)$$

has lower semicontinuity and/or convexity properties. A companion assumption is that the mapping

$$x \longrightarrow U(x)$$

is a measurable closed-valued multifunction [27]. In the latter case there exists a Borel measurable selector $\mu:S \longrightarrow C$ such that $\mu(x) \in U(x)$ for every state x (Kuratowski and Ryll-Nardzewski [18]). This is, of course, necessary if any policy is to exist at all.

The main fact regarding models of this type is that if g is lower semicontinuous, S and C are compact, and $x \longrightarrow U(x)$ is closed-valued and measurable, then the functions J_k defined by (1.4) and (1.5) are lower semicontinuous, the infimum in (1.5) is achieved for every x and k, and there are Borel measurable selectors $\mu_0,\ldots,\mu_{N-1}$ such that $\mu_k(x)$ achieves this infimum. The policy $(\mu_0,\ldots,\mu_{N-1})$ is optimal. This existence of an optimal policy is often an additional benefit of imposing topological conditions to insure that the problem is well-defined. For results in this direction, see Maitra [21], Schael [31]-[33], and Freedman [13]. Part of this thesis will deal with assumptions of this nature, not in order to resolve measure theoretic questions, but rather to give easily verifiable conditions on the system equation model which guarantee the existence of an optimal policy. In Chapter 5 we will show that these conditions also guarantee convergence of the dynamic programming algorithm over an infinite horizon to the optimal cost function and that this algorithm can be used to generate an optimal stationary

policy. This result generalizes the work of Maitra [21] in that we need not assume the one-stage cost is bounded and the discount factor is less than one.

Of the above mentioned papers, the one most comparable to the part of our own work that utilizes semicontinuity assumptions is Schael [32], which reaches the same conclusions we do under conditions in some ways more general and in other ways more restrictive than our own. In contrast to that work, our development does not appeal to the Kuratowski-Ryll-Nardzewski selection theorem [18], and so we can consider a more general control constraint. In particular, the model with a continuous transition kernel, a positive definite quadratic cost function on a finite dimensional Euclidean space, and no control constraint fits our framework but not that of [32]. This model provides the motivation for our line of analysis and is discussed more specifically in the next section.

Continuity and compactness assumptions are integral to much of the work that has been done in stochastic programming. This work differs from our own in both its aims and its framework. First of all, in the usual stochastic programming model, the decision maker cannot influence the distribution of future states (see Olsen [23]-[25], Rockafellar and Wets [28],[29], and the references contained therein). Secondly, assumptions of convexity, lower semicontinuity or both are made on the cost function, the model is designed for the Kuratowski-Ryll-Nardzewski selection theorem, and the analysis is carried out in a finite dimensional Euclidean state space. All of this is for the purpose of overcoming measurability problems. Results are not readily generalizable beyond Euclidean spaces (Rockafellar [27]). The thrust of the work is toward convex programming type results, i.e. duality and Kuhn-Tucker conditions for optimality, and so very specific structure of control and even

state constraints is assumed and powerful results are obtained.

Our work, by admitting less structure, can be done in Euclidean spaces and immediately generalized to Borel spaces, all of which are Borel isomorphic to Borel subsets of [0,1]. The need for this flexibility will become apparent in Chapter 6 when we take conditional probabilities as the states of a model. The less structured model is, of course, consistent with our desire to present a general framework for stochastic decision theory. Our control constraint is of the form

$$U(x) = \{u: (x,u)\varepsilon\Gamma\} = \Gamma_x,$$

where Γ is an analytic subset of SC and $\Gamma_x \neq \emptyset$ for every x. It can be shown that if $x \rightarrow U(x)$ is a closed-valued measurable multifunction, then $\Gamma = \{(x,u): u \varepsilon U(x)\}$ is analytic, indeed Borel. (See [27], Theorem 1E for the case when the control space is R^n. The proof found there can be generalized.) Thus we have generalized this constraint to a case where U(x) need not be closed for each x.

III. <u>The Borel Models</u>

The Borel space framework was introduced by Blackwell [5] and further refined by Strauch, Dynkin, Juskevic, Hinderer and others [35, 12, 16, 21, 6, 13]. The state and control spaces S and C were assumed to be Borel spaces, and the functions defining the model were assumed to be Borel measurable. Initial efforts were directed toward proving the existence of "nice" optimal or nearly optimal policies in this framework. Policies were required to be Borel measurable.

Under these conditions it is possible to prove the universal measurability of the optimal cost function and the existence for every $\epsilon > 0$ and probability measure p on S of a p-ϵ-optimal policy (Strauch [35], Theorems 7.1 and 8.1). A p-ϵ-optimal policy is one which leads to a cost which differs from the optimal cost by less than ϵ for p-almost every initial state. Even over a finite horizon the optimal cost function need not be Borel measurable, and there need not exist an everywhere ϵ-optimal policy (Blackwell [5], Example 2). The difficulty arises from the inability to choose a Borel measurable function $\mu_k : S \rightarrow C$ which nearly achieves the infimum in (1.5) uniformly in x. The nonexistence of such a function interferes with the construction of optimal policies via the dynamic programming algorithm (1.4) and (1.5), since one must first determine at each stage the measure p with respect to which it is satisfactory to nearly achieve the infimum in (1.5) for p-almost every x. This is essentially the same problem encountered with (1.6). The difficulties in constructing nearly optimal policies over an infinite horizon are more acute. Furthermore, from an applications point of view a p-ϵ-optimal policy, even if it can be constructed, is a much less appealing object than an everywhere ϵ-optimal policy, since in many situations the distribution p is unknown or may change when the system is operated repetitively, in which case a new p-ϵ-optimal policy must be computed.

<u>The main qualitative result of this thesis is that if the class of admissible policies in the Borel model is enlarged to include all universally measurable policies, then the existence of everywhere ϵ-optimal policies can be assured and, if the infimum in the dynamic programming algorithm (1.5) is attained for every x and k, then an everywhere optimal policy exists</u>. Thus the notion of p-optimality can be dispensed with. The only other work of a similar nature is that of Blackwell, Freedman and Orkin [6], who extended the

class of admissible policies to those which are analytically measurable. Under our assumption of a nonpositive cost, they proved for any $\epsilon>0$ the existence of an everywhere ϵ-optimal policy which, at stage k, chooses control u_k dependent on the entire history $(x_0,u_0,\ldots,u_{k-1},x_k)$, i.e. has the form $\pi=(\mu_0,\ldots,\mu_{N-1})$ or $\pi=(\mu_0,\mu_1,\ldots)$, where μ_k is of the form (1.3). We prove in Corollary 3.2.2 that when universally measurable policies are allowed, then under the same assumption of a nonpositive cost, an ϵ-optimal semi-Markov policy exists, i.e. μ_k has the form $\mu_k(x_0,x_k)$. Thus the intermediate states and controls can be forgotten. We also provide an example to the effect that without further assumptions, the dependence on x_0 is necessary.

We would like to point out that while this thesis uses universally measurable policies to prove results which hold everywhere, one can obtain the former results which allow only Borel measurable policies and hold p-almost everywhere as corollaries. This follows from the following observation, whose proof we sketch shortly:

(1.7) If X and Y are Borel spaces, $p_0,p_1,\ldots$ is a sequence of probability measures on X, and μ is a universally measurable map from X to Y, then there is a Borel measurable map μ' from X to Y such that

$$\mu(x) = \mu'(x)$$

for p_k-almost every x, k=0,1,....

As an example of how this observation can be used to obtain p-almost everywhere existence results from ours, consider Theorem 5.12. It states in part that if $\epsilon>0$ and the discount factor α is less than one, then an ϵ-optimal nonrandomized stationary policy exists, i.e. a policy $\pi=(\mu,\mu,\ldots)$ where μ is a

universally measurable mapping from S to C. Given p_o on S, this policy generates a sequence of measures $p_o, p_1, \ldots$ on S, where p_k is the distribution of the k-th state when the initial state has distribution p_o and the policy π is used. Let $\mu': S \to C$ be Borel measurable and equal to μ for p_k-almost every x, $k=0,1,\ldots$. Let $\pi' = (\mu', \mu', \ldots)$. Then it can be shown that for p_o-almost every initial state, the cost corresponding to π' equals the cost corresponding to π, so π' is a p_o-ϵ-optimal nonrandomized stationary Borel measurable policy. The existence of such a π' is a new result. This type of argument can be applied to all the existence results of Chapters 3 and 5.

We now sketch a proof of (1.7). Assume first that Y is a Borel subset of $[0,1]$. Then for $r \in [0,1]$, r rational, the set

$$U(r) = \{x : \mu(x) \le r\}$$

is universally measurable. For every k, let $p_k^*[U(r)]$ be the outer measure of $U(r)$ with respect to k and let $B_{k1}, B_{k2}, \ldots$ be a decreasing sequence of sets containing $U(r)$ such that

$$p_k^*[U(r)] = p_k[\bigcap_{j=1}^{\infty} B_{kj}].$$

Let $B(r) = \bigcap_{k=1}^{\infty} \bigcap_{j=1}^{\infty} B_{kj}$. Then

$$p_k^*[U(r)] = p_k[B(r)], \quad k=0,1,\ldots,$$

and the argument of Lemma 2.1 applies.

If Y is an arbitrary Borel space, it is Borel isomorphic to a Borel subset of $[0,1]$ and (1.7) follows.

Section 3. Summary of results.

This section summarizes the remainder of the thesis in a chapter-by-chapter fashion. New theorems will be indicated. If a result is shown to hold for every initial state in our model, whereas it was previously known to hold for p-almost every initial state when policies were required to be Borel measurable, we will say the "everywhere nature" of the result is new. The logical order of the presentation is Appendix A, Appendix B, Appendix C, and then Chapters 2 through 6.

Chapter 2 and the appendices, independent of the stochastic decision models, present the pertinent mathematics. No genuinely new results are obtained but rather slight extensions of existing results to cover the cases at hand are proved. Section 1 of Chapter 2 collects most of the notation and conventions used. Section 2 states and proves some elementary facts about universally measurable extended real-valued functions and universally measurable stochastic kernels. In Section 3 it is shown that infimizing a bivariate lower semianalytic function over one variable results in a lower semianalytic function. Theorem 2.4 establishes that when an extended real-valued lower semianalytic function is integrated against a Borel measurable stochastic kernel, the resulting function is lower semianalytic. These two facts will be used to guarantee that the functions generated by the dynamic programming algorithm are lower semianalytic. Theorem 2.5 relates to measurable selection of extremals and generalizes a result of Brown and Purves [7] in that it applies to minimization of lower semianalytic functions rather than Borel measurable functions. The theory of the remaining chapters rests on this theorem. It guarantees that for $\epsilon>0$, universally measurable functions $\mu_k : S \rightarrow C$ exist such that if the infimum in (1.5) is achieved at x, then $\mu_k(x)$

achieves it, and otherwise $\mu_k(x)$ comes within ϵ of achieving it. Section 4 develops the analogous facts about lower semicontinuous functions, except that under conditions of lower semicontinuity and compactness the infimum in (1.5) is always achieved and the selector which achieves it can be chosen to be Borel measurable.

Chapter 3 is devoted to definition and analysis of the finite horizon stochastic decision model. Section 1 defines the model. Section 2 defines the dynamic programming algorithm in terms of the operator T and shows that the algorithm generates the optimal cost function (Theorem 3.2). We then use the algorithm to construct ϵ-optimal policies in Corollary 3.2.2 and a sequence of policies exhibiting $\{\epsilon_n\}$ dominated convergence to optimality in Corollary 3.2.3. These corollaries and Example 3.1 show that the strongest possible structural results hold. The everywhere nature of the (F^+) result is new and the (F^-) results are completely new.

As remarked before, our framework is designed so that when the infimum in the dynamic programming algorithm is attained for every x and k, an optimal policy exists. This is the statement of Theorem 3.3, which is new.

Theorem 3.4 gives a set of easily verifiable conditions which guarantee the existence of an optimal policy. Note that if $S=R^n$, $C=R^m$, $\Gamma=SC$, $g(x,u)=x'Qx+u'Ru$, where Q is a positive semidefinite matrix and R is a positive definite matrix, then by replacing C by its one point compactification and taking $\Gamma^j=\{(x,u):\ u'u\leq j\}$, conditions (3.16)-(3.18) are satisfied. This justifies our earlier remark concerning the motivation behind these assumptions.

Chapter 4 shows in detail how the stochastic decision model can be converted to a deterministic one. Although the concept is not new (see Witsenhausen [38]), to the author's knowledge this has not been done in this systematic manner before. The conversion is carried through only for the infinite horizon model, as it is not necessary for the development in Chapter 3. It is also done only under assumptions (P),(N), or (D), although the models make sense under conditions of summability similar to those of Chapter 3. The models (P) and (N) are fundamentally different, and as the analysis proceeds, it would be necessary to make stronger assumptions than summability in order to more clearly differentiate the two cases. This is worked out in some detail in [12].

The conversion of the stochastic model to the deterministic one plays a central role in the analysis of the infinite horizon problem in Chapter 5, where structural results are applied to the deterministic model and then transferred to the stochastic model. This line of analysis is economical and results in exceedingly simple proofs of some otherwise difficult theorems. One such is given as Theorem 7.1 of Strauch [35] and here as Corollary 4.5.1, namely, that the optimal cost function across an infinite horizon is lower semianalytic. Another such result is the validity of the optimality equation given as Theorem 8.2 in Strauch [35] and as Theorem 5.1 here. The analysis also shows how results for stochastic models with measurability restrictions on the set of admissible policies can be obtained from general results on abstract dynamic programming models based on monotone mappings such as those of Denardo [8] and Bertsekas [4].

Chapter 5 begins with the well-known optimality equation for infinite horizon control and derives the similar functional equation for the cost corresponding to a stationary policy. In Theorem 5.3, existing structural properties relating the optimal cost function J^* to the T operator are proved for the Borel model. This theorem is new, as is its companion Theorem 5.4. Theorem 5.4 enables us to establish the necessary and sufficient condition for a stationary policy to be optimal under (P) and (D) found in Theorem 5.5. This condition is also found in Theorem 5.3 of Schael [32]. This condition can be applied as a test of optimality only if a stationary policy is already given, but we extend it to a necessary and sufficient condition for an optimal nonrandomized stationary policy to exist in Corollary 5.5.1. The corollary is new; indeed it is not true if policies are restricted to be Borel measurable. Theorem 5.6 gives a new but less satisfying condition for a stationary policy to be optimal under (N). An extension such as was done in Corollary 5.5.1 is not possible under (N).

Theorem 5.7 states that under (N) and (D) the finite horizon optimal cost functions converge to the infinite horizon optimal cost function as the horizon tends to infinity, and in Theorem 5.8 sharp bounds on the rate of convergence are provided under (D). This convergence does not always occur under (P) (see Strauch [35], Example 6.1 or Bertsekas [3], Chapter 6, Problem 4). Conditions equivalent to this convergence are given in Theorem 5.9. Theorems 5.7 - 5.9 are new for the Borel model. In Theorem 5.10 it is shown that the compactness of certain level sets in the control space implies the equivalent conditions of Theorem 5.9 and the existence of a nonrandomized stationary optimal policy. As corollaries of this theorem, we see that if U(x) is finite for each x or if the continuity and compactness conditions imposed in Chapter 3 to guarantee existence of an optimal policy over a finite

horizon are satisfied, then the conclusion of Theorem 5.10 holds. Theorem 5.11 indicates how, under these conditions, a nonrandomized stationary optimal policy can be obtained as a *limit of policies optimal over a finite horizon.*

Theorem 5.12 is the basic existence result for ϵ-optimal policies under (P) and (D). The everywhere nature of the existence of an ϵ-optimal nonrandomized Markov policy is new. The remainder of the theorem is completely new. Theorem 5.13 deals with case (N). This theorem is new both structurally and measure theoretically, the strongest previous result in this area being contained in Blackwell, Freedman and Orkin [6]. An interesting result in this area is the following due to Frid [14]: If under (N), the optimal value function J^* is everywhere finite, $0<\lambda<1$ and p is a probability measure on S, then a nonrandomized stationary policy π exists satisfying for p-almost every x

$$J_\pi(x) \leq \lambda J^*(x).$$

In particular, if J^* is bounded and $\epsilon>0$, a nonrandomized stationary p-ϵ-optimal policy exists. We have not been able to establish an everywhere ϵ-optimal version of this result.

Chapter 6 shows how more general models than that considered thus far can be reduced to our framework. Section 1 sketches this reduction for the nonstationary model. The nonstationary optimality equation is given as an example of the operation of this reduction. We use the nonstationary model in Theorem 6.2 to prove that for fixed p and $\epsilon>0$, a weakly p-ϵ-optimal nonrandomized Markov policy always exists. (Our "weak p-ϵ-optimality" is the "p-ϵ-optimality" considered by Hinderer [16].)

Section 2 accomplishes a similar reduction for the model of imperfect state information. A statistic sufficient for control is defined, and the values taken by this statistic become the states in the perfect state information model. Correspondences of costs of policies and optimal costs are established in Theorems 6.3 and 6.4. A discussion of the finite horizon dynamic programming algorithm for the imperfect state information model illustrates the reduction. The remaining theorems show that the identity mappings on the information vectors constitute a statistic sufficient for control, as do the mappings of the information vectors into the conditional distributions of the state.

Our definition of statistic sufficient for control is that given by Striebel [36] specialized to our framework, and our reduction of the imperfect state information model is similar to hers. Our work differs from Striebel's in that the existence of a statistic sufficient for control is guaranteed by our assumptions and, once this existence is shown, the entire theory of the previous chapters can be brought to bear on the imperfect state information model.

Appendix A presents, mostly without proof, the properties of analytic sets needed for the thesis.

Appendix B develops properties of the Borel space $P(X)$ of probability measures on a Borel space X. Theorem B.3 characterizes the Borel σ-algebra in $P(X)$ independent of the topology on $P(X)$. Such a theorem for compact X is available in the literature [9, Proposition 3.1] and has been used for noncompact X (Strauch [35], Blackwell, Freedman and Orkin [6]), the authors evidently intending an extension of the compact result by using Urysohn's Theorem to embed X in a compact metric space. The details of this development

have been carried through by Hinderer [16, Theorem 12.13], while we give an alternate proof. In Theorem B.8, the analyticity of $\{p \in P(X): P(A) > c\}$ when X is compact and A is analytic is shown to hold for noncompact X by appealing to Urysohn's Theorem. This result has also been used previously (Blackwell, Freedman and Orkin [6]).

Appendix C defines and establishes the existence of stochastic kernels in Borel spaces. The extension of Theorem C.1 in Theorem C.2 to include a measurable dependence on a parameter is crucial for the development of the filtering algorithm in Section 2 of Chapter 6. Theorem C.3 characterizes stochastic kernels as measurable maps from conditioning variable to probability measure.

CHAPTER 2

MATHEMATICAL PRELIMINARIES

Section 1. Notation

Let Y be a topological space. The space Y is *complete* if there is a metric d on Y consistent with its topology such that (Y,d) is a complete metric space. Thus any topological space homeomorphic to a complete space is itself complete.

If Y is a complete sepa[redacted] (topological) space, we denote by $\mathcal{B}_Y$ the smallest σ-algebra containing the open sets in Y. The sets in $\mathcal{B}_Y$ are called the *Borel subsets* of Y. We will often write $\underline{Y}$ to indicate a set in $\mathcal{B}_Y$.

Definition 2.1 If X is a Borel subset of some complete separable space Y, we say X is a *Borel space*. We understand X to have the relative topology.

It usually suits our purposes to treat Borel spaces without specifying the complete space in which they are embedded or the metric which makes that space complete. By definition a Borel space X is a metrizable topological space, but no particular metric is specified. The class of Borel subsets of a Borel space X (denoted $\mathcal{B}_X$) is the smallest σ-algebra containing the relatively open sets in X. The Borel subsets of X are also Borel spaces in the sense of Definition 2.1.

If X is a Borel space, $P(X)$ is the set of probability measures on $(X,\mathcal{B}_X)$ and is a Borel space in its own right (Appendix B). The set of bounded, continuous, real-valued functions on X will be denoted by $C(X)$. If d is a metric on X consistent with its topology, then $U_d(X)$ is the set of bounded, real-valued functions on X which are uniformly continuous with respect to d.

The letter R represents the real line. The symbol R^* represents the real line with $+\infty$ and $-\infty$ adjoined. Defining open neighborhoods of $+\infty$ as $(c,+\infty]$ and of $-\infty$ as $[-\infty,c)$, R^* becomes a complete separable space containing R as a Borel subset. The set of rational numbers is denoted by Q, and Q^* is defined analogous to R^*.

We follow the usual conventions with regard to ordering and arithmetic in R^*, with the exception that whenever the sums $+\infty-\infty$ and $-\infty+\infty$ occur, we set them equal to $+\infty$. If X is a set and $f:X\rightarrow R^*$, then $f^+(x)=\max\{0,f(x)\}$ and $f^-(x)=\max\{0,-f(x)\}$. If (X,d) is a metric space and $x\in X$, $A\subset X$, then $d(x,A)=\inf_{y\in A}d(x,y)$. If f and p are a measurable extended real-valued function and a probability measure on a space, respectively, we define

$$\int f\,dp = \int f^+dp - \int f^-dp. \tag{2.1}$$

Under our conventions, this sum is always defined. The integral may not be linear, but it always holds that

$$\int (f+g)dp \leq \int f\,dp + \int g\,dp. \tag{2.2}$$

If $\int f\,dp$ and $\int g\,dp$ are not infinite of different sign, then, of course, equality holds in (2.2).

If $B\subset X$, the function χ_B is defined to be identically one on B and zero otherwise. The symmetric difference operator Δ is defined by $A\Delta B = (A-B)\cup(B-A)$. The juxtaposition of two or more sets represents their cross product. For example, XY is the cross product of X and Y. Assuming X and Y are Borel spaces and letting $\mathcal{B}_X\mathcal{B}_Y$ be the product σ-algebra, by [26], Chapter 1, Theorem 1.10, $\mathcal{B}_X\mathcal{B}_Y$ is equal to $\mathcal{B}_{XY}$. If a product space XY is given, proj_X is the projection mapping onto the X-axis. If $D\subset XY$, D_x is the cross section

$D_x = \{y \varepsilon Y: (x,y) \varepsilon D\}$.

The countable cross product of the set of positive integers is denoted by N. We understand N to have the product topology, where the set of positive integers has the discrete topology. Thus defined, N is a Borel space. The space of irrationals in $(0,1)$ with the usual topology is denoted by N' and is homeomorphic to N [19, Section 3(IX)].

Section 2. Universally measurable functions

In this section we list for reference several known properties of universally measurable functions. Proofs are given for the convenience of the reader. For a definition of analytic sets and a short exposition of their properties, see Appendix A. The symbols $\mathcal{U}_X$ and $\mathcal{A}_X$ will denote the universal and analytic σ-algebras, respectively, of a Borel space X (Definitions A.5 and A.6).

Definition 2.2 Let X and Y be Borel spaces and f a function mapping $D \varepsilon \mathcal{U}_X$ into Y. If $f^{-1}(B) \varepsilon \mathcal{U}_X$ for every $B \varepsilon \mathcal{B}_Y$, then f is universally measurable. If $D \varepsilon \mathcal{A}_X$ and $f^{-1}(B) \varepsilon \mathcal{A}_X$ for every $B \varepsilon \mathcal{B}_Y$, then f is analytically measurable.

In a sense, the class of universally measurable extended real-valued functions is as large a class as we dare consider. If an extended real-valued function f is not in this class, then there is some probability measure $p \varepsilon P(X)$ such that the integral of f with respect to p cannot be defined without resort to p outer measure. If f is in this class, then $\int f\, dp$ is defined by (2.1) and the remarks following Definition A.5 and, provided we take care with the addition of infinities, all the classical integration theorems are at our disposal.

<u>Lemma 2.1</u> Let X be a Borel space and $f:X \to R^*$. The function f is universally measurable if and only if for every $p \in P(X)$ there is a Borel measurable $f_p:X \to R^*$ such that for p-almost every x, $f(x)=f_p(x)$.

Proof:

Suppose f is universally measurable and let $p \in P(X)$ be given. For $r \in Q^*$, let $U(r) = \{x: f(x) \leq r\}$. Then $f(x) = \inf \{r \in Q^*: x \in U(r)\}$. Let $B(r) \in \mathcal{B}_X$ be such that $p[B(r) \Delta U(r)] = 0$. Define

$$f_p(x) = \inf \{r \in Q^*: x \in B(r)\} = \inf_{r \in Q^*} r\Psi_{B(r)}(x),$$

where $\Psi_{B(r)}=1$ if $x \in B(r)$ and $\Psi_{B(r)}=+\infty$ otherwise. Then $f_p:X \to R^*$ is Borel measurable, and

$$\{x: f(x) \neq f_p(x)\} \subset \bigcup_{r \in Q} [B(r) \Delta U(r)]$$

has p-measure zero.

Conversely, if given $p \in P(X)$, there is a Borel measurable f_p such that $f=f_p$ p-almost everywhere, then

$$p(\{x: f(x) \leq c\} \Delta \{x: f_p(x) \leq c\}) = 0$$

for every $c \in R^*$, and the universal measurability of f follows. QED

Lemma 2.1 can be used to give an equivalent definition of $\int f\, dp$ when f is a universally measurable extended real-valued function on a Borel space X and $p \in P(X)$. Letting f_p be as above, we could define

$$\int f\, dp = \int f_p dp.$$

<u>Lemma 2.2</u> Let X and Y be Borel spaces and $q(dy|x)$ a universally measurable stochastic kernel on Y given X (see Appendix C). Then given $p \in P(X)$, there is

a Borel measurable stochastic kernel $q_p(dy|x)$ such that for p-almost every x, $q(\cdot|x)=q_p(\cdot|x)$.

Proof:

Since Y is separable and metrizable, the topology in Y can be generated by a countable basis of open neighborhoods $\mathcal{G} = \{G_1, G_2, \ldots\}$. Therefore $\mathcal{G}$ generates $\mathcal{B}_Y$. Let $\mathcal{F}$ be the class of sets in $\mathcal{G}$ and their finite intersections. For $F \in \mathcal{F}$, let f_F be a Borel measurable function for which

$$f_F(x) = q(F|x), \quad x \in B_F,$$

where B_F is a Borel measurable set with p-measure one. Such an f_F and B_F exist by Lemma 2.1. For $x \in \bigcap_{F \in \mathcal{F}} B_F$, let $q_p(\cdot|x)=q(\cdot|x)$. For $x \notin \bigcap_{F \in \mathcal{F}} B_F$, let $q_p(\cdot|x)$ be some fixed probability measure in $P(Y)$. The class of sets $\underline{Y}$ in $\mathcal{B}_Y$ for which $q_p(\underline{Y}|x)$ is Borel measurable in x is a Dynkin system containing $\mathcal{F}$. The class $\mathcal{F}$ is closed under finite intersections and generates $\mathcal{B}_Y$. The lemma follows from the Dynkin system theorem [1, Theorem 4.1.2]. QED

<u>Theorem 2.1</u> Let X and Y be Borel spaces and let $f: XY \to R^*$ be universally measurable. Let $q(dy|x)$ be a universally measurable stochastic kernel on Y given X. Then the mapping

$$x \to \int f(x,y) q(dy|x)$$

is universally measurable from X to R^*.

Proof:

Given $p \in P(X)$, there is a Borel measurable stochastic kernel $q_p(dy|x)$ such that $q_p(\cdot|x)=q(\cdot|x)$ for p-almost every x. Define a measure r on XY by specifying it on measurable rectangles to be [1, Theorem 2.6.2]

$$r(\underline{XY}) = \int_{\underline{X}} q_p(\underline{Y}|x)p(dx).$$

Let f_p be a Borel measurable function such that

$$f_p(x,y) = f(x,y)$$

for r-almost every (x,y). Then for every $E \in \mathcal{B}_X$,

$$\int_E \int_Y f_p(x,y)q_p(dy|x)p(dx) = \int_E \int_Y f(x,y)q_p(dy|x)p(dx).$$

This implies

$$\int_Y f_p(x,y)q_p(dy|x) = \int_Y f(x,y)q_p(dy|x) = \int_Y f(x,y)q(dy|x)$$

for p-almost every x. The left hand side is Borel measurable by (2.1), Corollary B.3.1 and Theorem C.3, so the right hand side is universally measurable by Lemma 2.1. QED

<u>Theorem 2.2</u> Let X and Y be Borel spaces, D a universally measurable subspace of X, and $f:D \to Y$ a universally measurable function. If $U \subset Y$ is universally measurable, then $f^{-1}(U)$ is universally measurable.

Proof:

Let m be a finite measure on X and define a measure m' on Y by

$$m'(B) = m[f^{-1}(B)],\ B \in \mathcal{B}_Y.$$

Since $U \in \mathcal{U}_Y$, there exists a Borel set $B \subset Y$ such that

$$m[f^{-1}(U) \Delta f^{-1}(B)] = m'[U \Delta B] = 0.$$

The set $f^{-1}(B)$ is in $\mathcal{U}_X$, and so there exists a Borel set $C \subset X$ such that $m[C \Delta f^{-1}(B)]=0$. Then

$$m[f^{-1}(U) \Delta C] = 0,$$

so $f^{-1}(U)$ is in the completion of $\mathcal{B}_X$ with respect to m. QED

Corollary 2.2.1 Let X,Y and Z be Borel spaces and $f:X \to Y$, $g:Y \to Z$ universally measurable. Then $g \circ f$ is universally measurable.

Corollary 2.2.2 Let X,Y and Z be Borel spaces and $f:X \to Y$, $g:Y \to Z$ analytically measurable. Then $g \circ f$ is universally measurable.

One might speculate that under the hypotheses of Corollary 2.2.2, the composition $g \circ f$ is analytically measurable. This is apparently an open question. If it could be answered in the affirmative, then the selector in Theorem 2.5 could be taken to be analytically measurable and analytically measurable policies (Definition 3.3) would suffice for the analysis of Chapters 3 - 6.

Section 3. Lower semianalytic functions and the key selection theorem

Definition 2.3 Let X be a Borel space and $f:X \to R^*$. If $\{x: f(x)<c\}$ is analytic for every $c \in R$, f is said to be lower semianalytic.

Note that as a consequence of Theorem A.2, f is lower semianalytic if and only if $\{x: f(x)<c\}$ is analytic for every $c \in R^*$. Also $\{x: f(x)<c\}$ is analytic for every $c \in R^*$ if and only if $\{x: f(x) \leq c\}$ is. If f and g are lower semianalytic, then for $c \in R$, $\{x: f(x)+g(x)<c\} = \bigcup_{r \in Q} \{x: f(x)<r, g(x)<c-r\}$ is analytic, so f+g is lower semianalytic. This is true even if $f(x)+g(x)=+\infty-\infty$, which by convention we take to be $+\infty$. As shown by the next theorem, lower semianalytic functions can be characterized as those functions obtained by infimizing bivariate Borel functions in one of their variables.

More importantly, if a bivariate lower semianalytic function is infimized in one of its variables, the result is again lower semianalytic. It is this closure under infimization which makes the dynamic programming algorithm of the subsequent chapters possible.

Theorem 2.3 Let X and Y be Borel spaces and $f:XY \to R^*$ be lower semianalytic. Then the function $\inf_{y \in Y} f(x,y)$ mapping X into R^* is lower semianalytic. Conversely, any lower semianalytic $g:X \to R^*$ is of the form

$$g(x) = \inf_{z \in N} f(x,z),$$

where $f:XN \to R^*$ is Borel measurable.

Proof:

For the first part of the theorem, observe that for $c \in R$,

$$\{x: \inf_{y \in Y} f(x,y) < c\} = \mathrm{proj}_X\{(x,y): f(x,y) < c\}$$

is analytic by Corollary A.3.1.

For the second part of the theorem, let $g:X \to R^*$ be lower semianalytic. For $r \in Q$, let $A(r)=\{x: g(x)<r\}$. Then $A(r)$ is analytic and by Theorem A.7, $A(r) = \mathrm{proj}_X F(r)$, where $F(r)$ is a closed set in XN. Define

$$G(r) = \bigcup_{r' \leq r} F(r')$$

and

$$f(x,z) = \inf \{r: (x,z) \in G(r)\} = \inf_{r \in Q} r\Psi_{G(r)}(x,z),$$

where $\Psi_{G(r)}(x,z)=1$ if $(x,z) \in G(r)$ and $\Psi_{G(r)}(x,z)=+\infty$ otherwise. The function f is Borel measurable.

If $g(x)<c$ for some $c \in R$, then there exists $r \in Q$ for which $g(x)<r<c$, and so $x \in A(r)$. There exists $z \in N$ such that $(x,z) \in G(r)$, and consequently $\inf_{z \in N} f(x,z) \leq r<c$. This shows $\inf_{z \in N} f(x,z)$ cannot be greater than $g(x)$.

If $\inf_{z \in N} f(x,z)<c$ for some $c \in R$, then there exists $r \in Q$ for which $\inf_{z \in N} f(x,z)<r<c$ and $(x,z) \in G(r)$. Thus for some $r' \in Q$, $r' \leq r$, we have $(x,z) \in F(r')$ and $x \in A(r')$. This implies $g(x)<r' \leq r<c$, which shows $g(x)$ cannot be greater than $\inf_{z \in N} f(x,z)$. QED

Theorem 2.4 Let X and Y be Borel spaces, $f: XY \rightarrow R^*$ lower semianalytic, and $q(dy|x)$ a Borel measurable stochastic kernel on Y given X. Then the function

$$x \rightarrow \int f(x,y)q(dy|x)$$

is lower semianalytic.

Proof:

Suppose $f \geq 0$. Let $f_n(x,y)=\min\{n, f(x,y)\}$. Then each f_n is lower semianalytic and $f_n \uparrow f$.

The set

$$E_n = \{(x,y,b): \ f_n(x,y) \leq b \leq n\}$$

$$= \bigcap_{k=1}^{\infty} \bigcup_{r \in Q} \{(x,y,b): \ f_n(x,y) < r, \ r \leq b + 1/k \leq n + 1/k\}$$

is analytic in XYR by Theorem A.2. Let λ be Lebesgue measure on R, $p \in P(XY)$ and $p\lambda$ the product measure on XYR. By Fubini's Theorem,

$$(p\lambda)(E_n) = \int_{XY} \int_R \chi_{E_n} d\lambda \, dp = \int_{XY} [n-f_n(x,y)]dp$$

$$= n - \int_{XY} f_n(x,y)dp.$$

For $c \in R$,

$$\{p \in P(XY): \int f(x,y)dp \leq c\} = \bigcap_{n=1}^{\infty} \{p \in P(XY): \int_{XY} f_n(x,y)dp \leq c\}$$

$$= \bigcap_{n=1}^{\infty} \{p \in P(XY): (p\lambda)(E_n) \geq n-c\}.$$

Define the mappings

$$\sigma: x \longrightarrow q(\cdot|x),$$

$$\tau_n: p \longrightarrow p\lambda_n,$$

where λ_n is λ restricted to $[0,n]$. By Theorem C.3, σ is Borel measurable. The mappings

$$p \longrightarrow p(A)\lambda_n(\underline{R}),\ A \in \mathcal{B}_{XY},\ \underline{R} \in \mathcal{B}_{[0,n]}$$

are Borel measurable, so the τ_n mappings are also (Corollary B.3.3).[1] The set

$$\{x: \int f(x,y)q(dy|x) \leq c\} = \bigcap_{n=1}^{\infty} \sigma^{-1}\tau_n^{-1}\{q \in P(XY[0,n]): q(E_n) \geq n-c\}$$

is analytic by Theorems A.2, A.3 and B.6.

Suppose $f \leq 0$. Let $f_n(x,y)=\max\{-n,f(x,y)\}$. Then each f_n is lower semianalytic and $f_n \downarrow f$. The sets $E_n = \{(x,y,b): f_n(x,y) \leq b \leq 0\}$ are analytic and

$$(p\lambda)(E_n) = \int_{XY}\int_R \chi_{E_n} d\lambda\, dp = -\int_{XY} f_n(x,y)dp.$$

For $c \in R$,

$$\{p \in P(XY): \int f(x,y)dp < c\} = \bigcup_{n=1}^{\infty} \{p \in P(XY): \int_{XY} f_n(x,y)dp < c\}$$

[1] Lebesgue measure on $[0,n]$ is not a probability measure, but the extension of Corollary B.3.3 to finite measures is immediate.

$$= \bigcup_{n=1}^{\infty} \{p \in P(XY): (p\lambda)(E_n) > -c\}.$$

Proceed as before.

In the general case,

$$\int f(x,y)q(dy|x) = \int f^{+}(x,y)q(dy|x) - \int f^{-}(x,y)q(dy|x).$$

The functions f^{+} and $-f^{-}$ are lower semianalytic, so by the above arguments, each of the summands on the right is lower semianalytic. The theorem follows from the remark following Definition 2.3. QED

Corollary 2.4.1 Let X be a Borel space, and let $f: X \rightarrow R^*$ be lower semianalytic. Then the function $p \rightarrow \int f\, dp$ is lower semianalytic on $P(X)$.

Proof:

Define a stochastic kernel on X given $P(X)$ by $q(\cdot|p)=p$. Apply Theorem 2.4. QED

We state and prove the key selection theorem. This is an extension of a theorem by L. D. Brown and R. Purves [7, Theorem 2], in that we allow f to be lower semianalytic rather than Borel measurable. Our proof parallels theirs.

Theorem 2.5 Let X and Y be Borel spaces, $D \subset XY$ an analytic set, and $f: D \rightarrow R^*$ lower semianalytic. Then

(a) The set

$$I = \{x \in \mathrm{proj}_X D: \text{ for some } y_o \in Y,\ f(x,y_o) = \inf_{y \in Y} f(x,y)\}$$

is universally measurable;

(b) For each $\epsilon > 0$, there is a universally measurable selector $\varphi: \mathrm{proj}_X D \rightarrow Y$ satisfying

$$f(x,\varphi(x)) = \min_{y\in Y} f(x,y) \text{ if } x\in I;$$

$$\leq \epsilon + \inf_{y\in Y} f(x,y) \text{ if } x\notin I,\ \inf_{y\in Y} f(x,y) \neq -\infty;$$

$$\leq -1/\epsilon \text{ if } x\notin I,\ \inf_{y\in Y} f(x,y) = -\infty.$$

Proof:

Assume first that $D=XY$. The set

$$E = \{(x,y,b):\ f(x,y) \leq b\}$$

$$= \bigcap_{k=1}^{\infty} \bigcup_{r\in Q^*} \{(x,y,b):\ f(x,y) \leq r,\ r \leq b + 1/k\}$$

is analytic in XYR^* by Theorem A.2. The set

$$A = \mathrm{proj}_{XR^*}(E)$$

is analytic in XR^* by Corollary A.3.1.

By von Neumann's Lemma (Theorem A.6), there is an analytically measurable $\rho: A \to Y$ such that $(x,\rho(x,b),b)\in E$ for every $(x,b)\in A$. Define $\psi: I \to Y$ by

$$\psi(x) = \rho(x, \inf_{y\in Y} f(x,y)).$$

For $x\in I$, $(x,\inf_{y\in Y} f(x,y))\in A$, so ψ is defined. The mapping

$$T: x \to (x, \inf_{y\in Y} f(x,y))$$

is analytically measurable (Theorem 2.3), and so ψ is universally measurable (Corollary 2.2.2), provided I is universally measurable. By Theorem 2.2,

$$I = \{x:\ (x,\inf_{y\in Y} f(x,y))\in A\} = T^{-1}(A)$$

is universally measurable.

For $\epsilon>0$, define the analytic sets

$$E_\epsilon = \{(x,y,b): f(x,y) \leq b + \epsilon\},$$
$$A_\epsilon = \mathrm{proj}_{XR^*} E_\epsilon.$$

Let $\rho_\epsilon : A_\epsilon \to Y$ be analytically measurable and satisfy $(x,\rho_\epsilon(x,b),b) \epsilon E_\epsilon$ for every $(x,b) \epsilon A_\epsilon$. Define $g: X \to R^*$ by

$$g(x) = \inf_{y \epsilon Y} f(x,y) \text{ if } \inf_{y \epsilon Y} f(x,y) \neq -\infty;$$
$$= -(1/\epsilon + \epsilon) \text{ otherwise.}$$

Then g is analytically measurable, so

$$\Psi_\epsilon(x) = \rho_\epsilon(x, g(x))$$

is universally measurable.

The function Φ defined by

$$\Phi(x) = \Psi(x) \text{ if } x \epsilon I;$$
$$= \Psi_\epsilon(x) \text{ if } x \notin I;$$

is universally measurable. For $x \epsilon I$,

$$f(x,\Phi(x)) = f(x,\Psi(x)) = f(x,\rho(x,\inf_{y \epsilon Y} f(x,y))) \leq \inf_{y \epsilon Y} f(x,y).$$

For $x \notin I$,

$$f(x,\Phi(x)) = f(x, \Psi_\epsilon(x)) = f(x,\rho_\epsilon(x,g(x)))$$
$$\leq \epsilon + g(x) \leq \epsilon + \inf_{y \epsilon Y} f(x,y) \text{ if } \inf_{y \epsilon Y} f(x,y) \neq -\infty;$$
$$\leq -1/\epsilon \text{ if } \inf_{y \epsilon Y} f(x,y) = -\infty.$$

Now if D is a proper subset of XY, extend f to XY by setting it equal to $+\infty$ outside D. Let Φ_1 be the selector given by the above argument applied to this extended f. Let $\Phi_2:\mathrm{proj}_X D \to Y$ be an analytically measurable function such that $(x,\Phi_2(x))\in D$ for every $x\in \mathrm{proj}_X D$. Set Φ equal to Φ_1 on $\{x\in \mathrm{proj}_X D: \inf_{y\in Y} f(x,y)<+\infty\}$ and equal to Φ_2 on $\{x\in \mathrm{proj}_X D: \inf_{y\in Y} f(x,y)=+\infty\}$. Then Φ has the required properties. QED.

Section 4. Lower semicontinuous functions and selection under compactness assumptions

In this section some known results on semicontinuous functions are listed. Most of the proofs are straight-forward or easily referenced and thus omitted.

Definition 2.4 Let X be a locally compact metric space. A subset K of X is σ-compact if $K=\bigcup_{j=1}^{\infty} K_j$, where each K_j is compact.

Lemma 2.3 Let X be a Borel space, Y a compact metric space and K_j, $j=1,2,\ldots$ a sequence of closed subsets of XY. Let $K=\bigcup_{j=1}^{\infty} K_j$. Then $\mathrm{proj}_X K$ is a Borel set and there is a Borel measurable map $\Phi:\mathrm{proj}_X K\to Y$ such that $(x,\Phi(x))\in K$ for all $x\in \mathrm{proj}_X K$.

Proof:

For each $x\in X$, K_x^j is closed in Y and consequently compact. Therefore $K_x=\bigcup_{j=1}^{\infty} K_x^j$ is σ-compact. The lemma follows from [7], Theorem 1. QED

Definition 2.5 Let X be a metric space and $f:X\to R^*$. The function f is lower semicontinuous if and only if $\{x: f(x)\le c\}$ is closed in X for every $c\in R$. Equivalently,

$$\liminf_{x \to x_0} f(x) \geq f(x_0)$$

for every $x_0 \in X$. If $-f$ is lower semicontinuous, f is said to be <u>upper semicontinuous</u>.

<u>Lemma</u> <u>2.4</u> Let X be a metric space, and let $f: X \to R^*$ be bounded below and lower semicontinuous. Then there exists a sequence $\{f_k\} \subset C(X)$ such that $f_k \uparrow f$.

Proof:

See [15], Section 42. QED

The following lemma lists two other properties of lower semicontinuous functions we will need.

<u>Lemma</u> <u>2.5</u> Let X and Y be metric spaces.

(a) Let f_1 and f_2 mapping X into R^* be lower semicontinuous and either both bounded above or both bounded below. Then f_1+f_2 is lower semicontinuous.

(b) Let $\{f_k\}$ be a sequence of lower semicontinuous functions mapping X into R^*. Then $\sup_k f_k$ is lower semicontinuous.

<u>Theorem</u> <u>2.6</u> Let X be a metric space, Y a compact metric space and $f: XY \to R^*$ lower semicontinuous. Then the function

$$g(x) = \min_{y \in Y} f(x,y)$$

is lower semicontinuous.

Theorem 2.6 establishes that on compact spaces, lower semicontinuous functions are closed under infimization in the same way that lower semianalytic functions are on Borel spaces (Theorem 2.3). There is also a selection theorem for lower semicontinuous functions (cf. Theorems 2.5 and

2.7). Similar to Theorem 2.4 for lower semianalytic functions and Borel measurable stochastic kernels, we will establish Theorem 2.8 for lower semicontinuous functions and continuous stochastic kernels.

<u>Theorem 2.7</u> Let X be a Borel space, Y a compact metric space and $f:XY \to R^*$ lower semicontinuous. Then there is a Borel measurable map $\varphi:X \to Y$ such that

$$f(x,\varphi(x)) = \min_{y \in Y} f(x,y)$$

for all $x \in X$.

Proof:

If f is real-valued, this theorem is analogous to a result given by Dubins and Savage [10, Chapter 2.16] and repeated by Maitra [21] for upper semicontinuous functions. The extension to extended real-valued functions is immediate. QED

The remainder of the chapter borrows from Schael [33].

<u>Lemma 2.6</u> Let X and Y be Borel spaces. For $p \in P(X)$, $q \in P(Y)$, let pq be the product measure on XY. The mapping $(p,q) \to pq$ is continuous.

Proof:

By Urysohn's Theorem [11, Chapter IX, Corollary 9.2], X and Y can be homeomorphically embedded in compact metric spaces $\hat{X}$ and $\hat{Y}$. For simplicity of notation, we treat X as a subset of $\hat{X}$ and Y as a subset of $\hat{Y}$. By Theorem B.4, X and Y are Borel subsets of $\hat{X}$ and $\hat{Y}$ respectively.

Let d be a metric on XY consistent with its topology and $f \in U_d(XY)$. Extend f to $\hat{f} \in C(\hat{X}\hat{Y})$ by Lemma B.1 and the Tietze extension theorem [11]. By the Stone-Weierstrass Theorem [34, Section 36, Theorem A], the set of finite

linear combinations of the form

$$\sum_{j=1}^{K} \hat{g}_j(x)\hat{h}_j(y), \quad \hat{g}_j \epsilon C(\hat{X}), \quad \hat{h}_j \epsilon C(\hat{Y}),$$

is dense in $C(\hat{X}\hat{Y})$ with respect to the supremum norm. Thus for given $\epsilon > 0$, it is possible to find such a linear combination $\sum_{j=1}^{K} \hat{g}_j(x)\hat{h}_j(y)$ which approximates $\hat{f}$ uniformly to within ϵ. The restrictions g_j and h_j of $\hat{g}_j$ and $\hat{h}_j$ to X and Y are in $C(X)$ and $C(Y)$ respectively. By Theorem B.1, if $p_n \to p$ in $P(X)$ and $q_n \to q$ in $P(Y)$,

$$\lim\sup_n \left| \int_{XY} f\, d(p_n q_n) - \int_{XY} f\, d(pq) \right| \leq \lim\sup_n \int_{XY} \left| f - \sum_{j=1}^{K} g_j h_j \right| d(p_n q_n)$$

$$+ \sum_{j=1}^{K} \lim_n \left| \int_X g_j dp_n \int_Y h_j dq_n - \int_X g_j dp \int_Y h_j dq \right| + \int_{XY} \left| \sum_{j=1}^{K} g_j h_j - f \right| d(pq) \leq 2\epsilon.$$

The continuity of $(p,q) \to pq$ follows from Theorem B.1. QED

<u>Definition 2.6</u> Let X and Y be Borel spaces and $q(dy|x)$ a stochastic kernel on Y given X. If the mapping $x \to q(dy|x)$ is continuous from X to $P(Y)$, the stochastic kernel q is said to be <u>continuous</u>.

<u>Lemma 2.7</u> Let X and Y be Borel spaces, $f \epsilon C(XY)$, and $q(dy|x)$ a continuous stochastic kernel on Y given X. Then the function

$$x \to \int f(x,y) q(dy|x)$$

is in $C(X)$.

Proof:

By Lemma 2.6 and Corollary B.7.1, the mapping

$$x \to p_x q(dy|x)$$

is continuous, where p_x is the probability measure assigning unit mass to the

point x. Compose this mapping with $r \to \int f\, dr$. QED

<u>Theorem</u> <u>2.6</u> Let X and Y be Borel spaces, $q(dy|x)$ a continuous stochastic kernel on Y given X, and $f: XY \to R^*$ bounded below and lower semicontinuous. Then the function

$$x \to \int f(x,y)\, q(dy|x)$$

is bounded below and lower semicontinuous.

Proof:

Use Lemma 2.4, 2.5(b) and 2.7. QED

CHAPTER 3

THE FINITE HORIZON MODEL

Section 1. The stochastic model

Definition 3.1 A stochastic decision model (SM) is the nine-tuple $(S,C,\Gamma,W,p,f,\alpha,g',N)$ described below. The letters x and u are used to denote elements of S and C respectively.

S: State space. A nonempty Borel space.

C: Control space. A nonempty Borel space.

Γ: Constraint set. An analytic subset of SC. For every $x \in S$, Γ_x is nonempty.

W: Disturbance space. A nonempty Borel space.

$p(dw|x,u)$: Disturbance kernel. A Borel measurable stochastic kernel on W given SC.

f: System function. Maps SCW into S and is measurable with respect to the product Borel σ-algebra.

α: Discount factor. $0 < \alpha \leq 1$.

g': One-stage cost function. Maps SCW into R^* and is lower semianalytic.

N: Horizon. A positive integer.

The system moves from state x_k to state x_{k+1} via the system equation

$$x_{k+1} = f(x_k,u_k,w_k), \quad k=0,1,\dots,N-1,$$

and incurs cost at each stage of $g'(x_k,u_k,w_k)$. The disturbances w_k are random objects with probability distributions $p(dw_k|x_k,u_k)$. The goal is to choose u_k dependent on the current state x_k so as to minimize

$$E\{\sum_{k=0}^{N-1}\alpha^k g'(x_k,u_k,w_k)\}.$$

We make this discussion precise with Definitions 3.2 - 3.6.

Definition 3.2 Given (SM) of Definition 3.1, the state transition kernel is the Borel measurable stochastic kernel on S given SC defined by

$$(3.1)\qquad t(A|x,u) = p(\{w: f(x,u,w)\varepsilon A\}|x,u) = p(f^{-1}(A)_{(x,u)}|x,u).$$

Thus defined, $t(A|x,u)$ is the probability that $x_{k+1}=f(x_k,u_k,w_k)$ is in A given that $(x_k,u_k)=(x,u)$. For fixed (x,u), $t(\cdot|x,u)$ is clearly a probability measure on S. To show that t is Borel measurable, we show that $p(B_{(x,u)}|x,u)$ is measurable for each Borel subset B of SCW. The sets B for which $p(B_{(x,u)}|x,u)$ is measurable form a Dynkin system, so by the Dynkin system theorem [1, Theorem 4.1.2], we need only verify that $p((\underline{SCW})_{(x,u)}|x,u)$ is measurable for $\underline{S}\varepsilon\mathcal{B}_S$, $\underline{C}\varepsilon\mathcal{B}_C$, $\underline{W}\varepsilon\mathcal{B}_W$. But

$$\begin{aligned} p((\underline{SCW})_{(x,u)}|x,u) &= p(\underline{W}|x,u) \text{ if } (x,u)\varepsilon\ \underline{SC}; \\ &= 0 \text{ otherwise.} \end{aligned}$$

Using p it is possible to "integrate out" the w in g' so that the disturbance space W disappears entirely from the model description. To do this, define the (reduced) one-stage cost function

$$(3.2)\qquad g(x,u) = \int_W g'(x,u,w)p(dw|x,u).$$

Then g is a lower semianalytic function on SC (Theorem 2.4).

Definition 3.3 A policy in (SM) is a sequence $\pi=(\mu_0,\mu_1,\ldots,\mu_{N-1})$ such that for each k, $\mu_k(du_k|x_0,u_0,\ldots,u_{k-1},x_k)$ is a universally measurable stochastic kernel on C given SC...CS satisfying

$$\mu_k(\Gamma_{x_k}|x_0,u_0,\ldots,u_{k-1},x_k) = 1$$

for every $(x_0,u_0,\ldots,u_{k-1},x_k)$. If for each k, μ_k is parameterized only by (x_0,x_k), π is a <u>semi-Markov policy</u>. If μ_k is parameterized only by x_k, π is a <u>Markov policy</u>. If for each k and $(x_0,u_0,\ldots,u_{k-1},x_k)$, $\mu_k(\cdot|x_0,u_0,\ldots,u_{k-1},x_k)$ assigns mass one to some point in C, π is <u>nonrandomized</u>. In this case, by a slight abuse of notation, π can be considered to be a sequence of universally measurable mappings $\mu_k:S\rightarrow C$ such that

$$(x_k,\mu_k(x_0,u_0,\ldots,u_{k-1},x_k))\in\Gamma$$

for every $(x_0,u_0,\ldots,u_{k-1},x_k)$. A policy is said to be <u>Borel measurable</u> if all its stochastic kernel components are.

We denote by Π' the set of all policies in (SM) and by Π the set of all Markov policies. We will show that in many cases it is not necessary to go outside Π to find the "best" available policy. In most cases, this "best" policy can be taken to be nonrandomized. By von Neumann's Lemma (Theorem A.6), there exists at least one nonrandomized Markov policy, so Π and Π' are nonempty.

For $p_0\in P(S)$ and $\pi=(\mu_0,\ldots,\mu_{N-1})\in\Pi'$, we say the measures $q_k\in P(SC)$, $k=0,1,\ldots,N-1$, are <u>generated from p_0 by π</u> if for every $\underline{S}_k\in\mathcal{B}_S$, $\underline{C}_k\in\mathcal{B}_C$,

$$q_k(\underline{S}_k\underline{C}_k) = \int_S\int_C\cdots\int_{\underline{S}_k}\mu_k(\underline{C}_k|x_0,u_0,\ldots,u_{k-1},x_k)\, t(dx_k|x_{k-1},u_{k-1})\cdots\mu_0(du_0|x_0)p_0(dx_0). \tag{3.3}$$

If $(q_0,\ldots,q_{N-1})$ is generated from p_x by π, we say that it is <u>generated from x by π</u>.

<u>Definition 3.4</u> If $x\in S$, $\pi\in\Pi'$, $K=1,\ldots,N$, and $(q_0,\ldots,q_{N-1})$ is generated by π from x, then the <u>K-stage cost function corresponding to π</u> at x is

$$J_{K,\pi}(x) = \sum_{k=0}^{K-1} \alpha^k \int g \, dq_k. \tag{3.4}$$

The <u>cost function</u> corresponding to π is $J_{N,\pi}$.

<u>Definition 3.5</u> Given $x \in S$, $K \leq N$, the <u>K-stage optimal cost function</u> at x is

$$J_K^*(x) = \inf_{\pi \in \Pi'} J_{K,\pi}(x).$$

The <u>optimal cost function</u> is J_N^*.

Note that J_K^* is independent of the horizon N as long as $K \leq N$.

<u>Definition 3.6</u> If $\epsilon > 0$, the policy π is <u>K-stage ϵ-optimal at x</u> provided

$$\begin{aligned} J_{K,\pi}(x) &\leq J_K^*(x) + \epsilon \text{ if } J_K^*(x) > -\infty; \\ &\leq -1/\epsilon \text{ if } J_K^*(x) = -\infty. \end{aligned}$$

If $J_{K,\pi}(x) = J_K^*(x)$, then π is <u>K-stage optimal at x</u>. If π is K-stage ϵ-optimal or K-stage optimal at every x, it is <u>K-stage ϵ-optimal</u> or <u>K-stage optimal</u> respectively. If $\{\epsilon_n\}$ is a sequence of positive numbers with $\epsilon_n \downarrow 0$, a sequence of policies $\{\pi_n\}$ exhibits $\{\epsilon_n\}$ <u>dominated convergence to optimality</u> provided

$$\lim_{n \to \infty} J_{K,\pi_n} = J_K^*,$$

and for n=2,3,...

$$\begin{aligned} J_{K,\pi_n}(x) &\leq J_K^*(x) + \epsilon_n \text{ if } J_K^*(x) > -\infty, \\ J_{K,\pi_n}(x) &\leq J_{K,\pi_{n-1}}(x) + \epsilon_n \text{ if } J_K^*(x) = -\infty. \end{aligned}$$

If K=N, we suppress the qualifier "K-stage" in the above terms.

If the model (SM) is such that for all $x \in S$ and $\pi \in \Pi'$, the measures $(q_0, \dots, q_{N-1})$ generated from x by π satisfy $\int g^+ \, dq_k < \infty$, $k=0,\dots,N-1$, we say (SM) is <u>summable above</u> and use the symbol (F^-) to show that a result holds under this assumption. If $\int g^- \, dq_k < \infty$ for every q_k which is an element of a sequence of measures generated from some $x \in S$ by some $\pi \in \Pi'$, we say (SM) is <u>summable below</u> and use the symbol (F^+) to show that a result holds under this assumption. Under (F^+), $J^*_{K,\pi}$ is a mapping from S to $(-\infty,+\infty]$ for each policy π and each K, while under (F^-), $J^*_{K,\pi}$ maps S into $[-\infty,+\infty)$.

It will often be convenient to subscript the state and control spaces as is done in the next theorem. Except for Chapter 6, Section 1, S_k will always be a copy of S and C_k will always be a copy of C.

<u>Theorem 3.1</u> If $x \in S$ and $\pi' \in \Pi'$, then there is a Borel measurable Markov policy π such that $J_{K,\pi'}(x) = J_{K,\pi}(x)$, $K=1,\dots,N$.

Proof:

Let $\pi' = (\mu_0', \dots, \mu_{N-1}')$. Given $x \in S$, let $(q_0', \dots, q_{N-1}')$ be the sequence of measures generated by π' from x. Let $\mu_k(du_k | x_k)$ be the Borel measurable stochastic kernel obtained by decomposing q_k' (Theorem C.1), i.e. for every $\underline{S}_k \in \mathcal{B}_S$, $\underline{C}_k \in \mathcal{B}_C$,

$$q_k'(\underline{S}_k \underline{C}_k) = \int_{\underline{S}_k} \mu_k(\underline{C}_k | x_k) q_k'(dx_k C_k), \quad k=0,\dots,N-1.$$

Let $\pi = (\mu_0, \dots, \mu_{N-1})$ and let $(q_0, \dots, q_{N-1})$ be the sequence of measures generated by π from x. It suffices to show $q_k = q_k'$, $k=0,\dots,N-1$. We proceed by induction.

For $k=0$, $\underline{S}_0 \in \mathcal{B}_S, \underline{C}_0 \in \mathcal{B}_C$,

$$q_0'(\underline{S}_0\underline{C}_0) = \int_{\underline{S}_0} \mu_0(\underline{C}_0|x_0)p_x(dx_0) = q_0(\underline{S}_0\underline{C}_0).$$

If $q_k=q_k'$, then for $\underline{S}_0 \in \mathcal{B}_S$, $\underline{C}_0 \in \mathcal{B}_C$,

$$q_{k+1}'(\underline{S}_{k+1}\underline{C}_{k+1}) = \int_{\underline{S}_{k+1}} \mu_{k+1}(\underline{C}_{k+1}|x_{k+1})q_{k+1}'(dx_{k+1}C_{k+1})$$

$$= \int_{S_k C_k}\int_{\underline{S}_{k+1}} \mu_{k+1}(\underline{C}_{k+1}|x_{k+1})t(dx_{k+1}|x_k,u_k)q_k'(d(x_k,u_k))$$

$$= q_{k+1}(\underline{S}_{k+1}\underline{C}_{k+1}),$$

where the last equality follows from the induction hypothesis. QED

Corollary 3.1.1 For every $k \le N$ and $x \in S$, $\inf_{\pi\in\Pi'} J_{k,\pi}(x)=\inf_{\pi\in\Pi} J_{k,\pi}(x)$.

Section 2. The dynamic programming operators. Existence of ϵ-optimal policies.

Let $U(C|S)$ denote the set of universally measurable stochastic kernels μ on C given S which satisfy $\mu(\Gamma_x|x)=1$ for every $x \in S$. Thus $\Pi=U(C|S)U(C|S)\ldots U(C|S)$, where there are N factors.

Definition 3.7 Let $J:S\to R^*$ be universally measurable and $\mu \in U(C|S)$. The operator T_μ mapping J into $T_\mu J:S\to R^*$ is defined by

$$(T_\mu J)(x) = \int_C [g(x,u) + \alpha\int_S J(x')t(dx'|x,u)]\mu(du|x)$$

for every $x \in S$.

By Theorem 2.1, $T_\mu J$ is universally measurable. We show in Lemma 3.2 that under (F^+) or (F^-), the cost corresponding to a policy $\pi=(\mu_0,\ldots,\mu_{N-1})$ can be defined in terms of the operators $T_{\mu_0},\ldots,T_{\mu_{N-1}}$.

<u>Definition 3.8</u> Let $J:S \to R^*$ be lower semianalytic. The <u>operator T</u> mapping J into $TJ:S \to R^*$ is defined by

$$(TJ)(x) = \inf_{u \in \Gamma_x}\{g(x,u) + \alpha \int J(x')t(dx'|x,u)\}$$

for every $x \in S$.

By Theorems 2.3 and 2.4, TJ is lower semianalytic. We show in Theorem 3.2 that under (F^+) or (F^-) the optimal cost can be defined in terms of the operator T.

<u>Lemma 3.1</u> Let $J:S \to R^*$ be lower semianalytic. Then for $\epsilon > 0$, there exists $\mu \in U(C|S)$ such that

$$\begin{aligned}(T_\mu J)(x) &\leq (TJ)(x) + \epsilon \text{ if } (TJ)(x) > -\infty;\\ &= -\infty \text{ if } (TJ)(x) = -\infty.\end{aligned}$$

Proof:

By Theorem 2.5, there are universally measurable selectors $\mu_m:S \to C$ such that for $m=1,2,\ldots$ and $x \in S$, $\mu_m(x) \in \Gamma_x$ and

$$\begin{aligned}(T_{\mu_m} J)(x) &\leq (TJ)(x) + \epsilon \text{ if } (TJ)(x) > -\infty;\\ &\leq -2^m \text{ if } (TJ)(x) = -\infty.\end{aligned}$$

Let $\mu(\cdot|x)$ assign mass one to $\mu_1(x)$ if $(TJ)(x) > -\infty$ and assign mass $1/2^m$ to $\mu_m(x)$, $m=1,2,\ldots$, if $(TJ)(x) = -\infty$. Then μ has the desired properties. QED

<u>Lemma 3.2</u> Let $\pi = (\mu_0,\ldots,\mu_{N-1})$ be in Π, $K=1,\ldots,N$, and J_0 identically zero. Then

$$J_{K,\pi} \geq (T_{\mu_0} \cdots T_{\mu_{K-1}})J_0. \tag{3.5}$$

Under (F^+) or (F^-), equality holds in (3.5).

Proof:

Let $\pi=(\mu_0,\ldots,\mu_{N-1})$ be in Π, $x\varepsilon S$ and $(q_0,\ldots,q_{N-1})$ be generated from x by π. Then for $K=1,\ldots,N$,

$$J_{K,\pi}(x) = \sum_{k=0}^{K-1}\alpha^k\int_{S_kC_k} g\, dq_k$$

$$= \sum_{k=0}^{K-2}\alpha^k\int_{S_kC_k} g\, dq_k$$

$$+\int_{S_{K-2}C_{K-2}}\alpha^{K-1}\int_{S_{K-1}}(T_{\mu_{K-1}}J_o)(x_{K-1})t(dx_{K-1}|x_{K-2},u_{K-2})q_{K-2}(d(x_{K-2},u_{K-2}))$$

$$= \sum_{k=0}^{K-3}\alpha^k\int_{S_kC_k} g\, dq_k + \int_{S_{K-3}C_{K-3}}\alpha^{K-2}\int_{S_{K-2}}\int_{C_{K-2}} g(x_{K-2},u_{K-2})$$

$$\mu_{K-2}(du_{K-2}|x_{K-2})t(dx_{K-2}|x_{K-2},u_{K-3})q_{K-3}(d(x_{K-3},u_{K-3}))$$

$$+\int_{S_{K-3}C_{K-3}}\alpha^{K-2}\int_{S_{K-2}}\int_{C_{K-2}}\alpha\int_{S_{K-1}}(T_{\mu_{K-1}}J_o)(x_{K-1})t(dx_{K-1}|x_{K-2},u_{K-2})$$

$$\mu_{K-2}(du_{K-2}|x_{K-2})t(dx_{K-2}|x_{K-3},u_{K-3})q_{K-3}(d(x_{K-3},u_{K-3}))$$

$$\geq \sum_{k=0}^{K-3}\alpha^k\int_{S_kC_k} g\, dq_k + \int_{S_{K-3}C_{K-3}}\alpha^{K-2}\int_{S_{K-2}}(T_{\mu_{K-2}}T_{\mu_{K-1}})(J_o)(x_{K-2})$$

$$t(dx_{K-2}|x_{K-3},u_{K-3})q_{K-3}(d(x_{K-3},u_{K-3})) \text{ by (2.2).}$$

Repeating this procedure finitely many times, we eventually obtain

$$J_{K,\pi}(x) \geq (T_{\mu_o}\cdots T_{\mu_{K-1}})(J_o)(x).$$

If (SM) is summable below or summable above, the above inequalities are equalities. QED

<u>Lemma 3.3</u> (F^+)

If J_o is identically zero, then $(T^kJ_o)(x)>-\infty$ for every $x\varepsilon S$, $k=1,\ldots,N$.

Proof:

Suppose for some $K\leq N$ and $\bar{x}\varepsilon S$ that

$$(T^j J_o)(x) > -\infty, \ j=0,\dots,K-1,$$

for every $x \in S$, and

$$(T^K J_o)(\bar{x}) = -\infty.$$

By Theorem 2.5, there are universally measurable selectors $\mu_j : S \to C$, $j=1,\dots,K$, such that $\mu_j(x) \in \Gamma_x$ and

$$T_{\mu_{K-j}}(T^{j-1}J_o)(x) \le (T^j J_o)(x) + 1, \ j=1,\dots,K-1,$$

for every $x \in S$. Then

$$\begin{aligned}(T_{\mu_1}\dots T_{\mu_{K-1}})(J_o) &\le (T_{\mu_o}\dots T_{\mu_{K-2}})(TJ_o + 1)\\ &\le (T_{\mu_o}\dots T_{\mu_{K-3}})(T^2 J_o + 1 + \alpha)\\ &\le T^{K-1}J_o + N,\end{aligned}$$

where the last inequality is obtained by repeating the process used to obtain the first two inequalities. By Lemma 3.1, there is a stochastic kernel $\mu_o \in U(C|S)$ such that

$$T_{\mu_o}(T^{K-1}J_o)(\bar{x}) = -\infty.$$

Then

$$\begin{aligned}(T_{\mu_o}T_{\mu_1}\dots T_{\mu_{K-1}})(J_o)(\bar{x}) &\le T_{\mu_o}(T^{K-1}J_o + N)(\bar{x})\\ &= -\infty.\end{aligned}$$

For $\mu \in U(C|S)$, let $\pi = (\mu_o,\dots,\mu_{K-1},\mu,\dots,\mu) \in \Pi$ and let $(q_o,\dots,q_{N-1})$ be generated from $\bar{x}$ by π. By Lemma 3.2,

$$\sum_{j=0}^{K-1} \alpha^K \int g \, dq_j = J_{K,\pi}(\bar{x}) = (T_{\mu_o}\dots T_{\mu_{K-1}})(J_o)(\bar{x}) = -\infty,$$

and so for some j, $\int g^- \, dq_j = \infty$. This contradicts the fact that (SM) is summable below. QED

Lemma 3.4 Let $\{J_k\}$ be a sequence of extended real-valued universally measurable functions on S and let μ be an element of $U(C|S)$.

(a) If $(T_\mu J_1)(x) < \infty$ for every $x \in S$ and $J_k \downarrow J$, then $T_\mu J_k \downarrow T_\mu J$.

(b) If $(T_\mu J_1^-)(x) < \infty$ for every $x \in S$, $g \geq 0$, and $J_k \uparrow J$, then $T_\mu J_k \uparrow T_\mu J$.

(c) If $\{J_k\}$ is uniformly bounded, g is bounded, and $J_k \rightarrow J$, then $T_\mu J_k \rightarrow T_\mu J$.

Proof:

Assume first that $T_\mu J_1 < \infty$ and $J_k \downarrow J$. Fix x. Since

$$\int [g(x,u) + \alpha \int J_1(x')t(dx'|x,u)]\mu(du|x) < \infty ,$$

we have

$$g(x,u) + \alpha \int J_1(x')t(dx'|x,u) < \infty$$

for $\mu(\cdot|x)$-almost all u. By the monotone convergence theorem,

$$g(x,u) + \alpha \int J_k(x')t(dx'|x,u) \downarrow g(x,u) + \alpha \int J(x')t(dx'|x,u)$$

for $\mu(\cdot|x)$-almost all u. Apply the monotone convergence theorem again to conclude $(T_\mu J_k)(x) \downarrow (T_\mu J)(x)$.

If $T_\mu J_1^- < \infty$, $g \geq 0$, and $J_k \uparrow J$, the same type of argument applies. If $\{J_k\}$ is uniformly bounded, g is bounded, and $J_k \rightarrow J$, a similar argument using the bounded convergence theorem applies. QED

Lemma 3.5 Let $\{J_k\}$ be a sequence of universally measurable functions from S to R^* and μ a universally measurable function from S to C whose graph lies in Γ. Suppose for some sequence $\{\epsilon_k\}$ of positive numbers with $\sum_{k=1}^{\infty} \epsilon_k < \infty$, we have

for every $x \in S$,

$$\int J_1^+(x')t(dx'|x,\mu(x)) < +\infty,$$

$$\lim_{k\to\infty} J_k = J,$$

and for k=2,3,...

$$J(x) \le J_k(x) \le J(x) + \epsilon_k \text{ if } J(x)>-\infty,$$

$$J_k(x) \le J_{k-1}(x) + \epsilon_k \text{ if } J(x)=-\infty.$$

Then

$$\lim_{k\to\infty} T_\mu J_k = T_\mu J.$$

Proof:

Since $J \le J_k$ for every k, it is clear that

$$T_\mu J \le \lim\inf_{k\to\infty} T_\mu J_k.$$

For $x \in S$,

$$\lim\sup_{k\to\infty} (T_\mu J_k)(x) \le g(x,\mu(x)) + \alpha \lim\sup_{k\to\infty} \int_{\{J>-\infty\}} J_k(x')t(dx'|x,\mu(x))$$

$$+ \lim\sup_{k\to\infty} \int_{\{J=-\infty\}} J_k(x')t(dx'|x,\mu(x)).$$

Now

$$\lim\sup_{k\to\infty} \int_{\{J>-\infty\}} J_k(x')t(dx'|x,\mu(x))$$

$$\le \lim\sup_{k\to\infty} [\int_{\{J>-\infty\}} J(x')t(dx'|x,\mu(x)) + \epsilon_k]$$

$$\le \int_{\{J>-\infty\}} J(x')t(dx'|x,\mu(x)).$$

If $J(x')=-\infty$, then

$$J_k(x') + \sum_{n=k+1}^{\infty} \epsilon_n \downarrow J(x'),$$

$t(dx'|x,\mu(x))$-almost surely, and since $\int [J_1^+(x')+\sum_{n=2}^{\infty}\epsilon_n] t(dx'|x,\mu(x))<\infty$,

$$\begin{aligned}
\limsup_{k\to\infty} & \int_{\{J=-\infty\}} J_k(x')t(dx'|x,\mu(x)) \\
& \leq \lim_{k\to\infty} \int_{\{J=-\infty\}} [J_k(x') + \sum_{n=k+1}^{\infty} \epsilon_n] t(dx'|x,\mu(x)) \\
& = \int_{\{J=-\infty\}} J(x')t(dx'|x,\mu(x)).
\end{aligned}$$

It follows that

$$\limsup_{k\to\infty} T_\mu J_k \leq T_\mu J. \quad \text{QED}$$

The dynamic programming algorithm over a finite horizon is executed by beginning with the identically zero function on S and applying the operator T successively N times. The next theorem says that this procedure generates the optimal cost function. In Corollary 3.2.2, we show how ϵ-optimal policies can also be obtained from this algorithm.

<u>Theorem 3.2</u> $(F^+)(F^-)$

Let J_o be the identically zero function on S. Then $J_N^*=T^N J_o$.

Proof:

For any $\pi=(\mu_o,\ldots,\mu_{N-1})\in\Pi$, $K\leq N$,

$$\begin{aligned}
J_{K,\pi} &= (T_{\mu_o}\cdots T_{\mu_{K-1}})(J_o) \\
&\geq (T_{\mu_o}\cdots T_{\mu_{K-2}})(TJ_o) \\
&\geq T^K J_o,
\end{aligned} \tag{3.6}$$

where the last inequality is obtained by repeating the process used to obtain the first inequality. Infimizing over $\pi\in\Pi$ when $K=N$, we obtain $J_N^*\geq T^N J_o$.

If (SM) is summable below, then by Lemma 3.3, $(T^k J_o) > -\infty$, $k=1,\ldots,N$. For $\epsilon > 0$, there are universally measurable selectors $\hat{\mu}_k : S \to C$, $k=0,\ldots,N-1$, with $\hat{\mu}_k(x) \epsilon \Gamma_x$ and

$$-\infty < T_{\hat{\mu}_{N-k}}(T^{k-1}J_o)(x) \leq (T^k J_o)(x) + \epsilon/N, \quad k=1,\ldots,N,$$

for every $x \epsilon S$ (Theorem 2.5). Then

$$(3.7) \qquad (T_{\hat{\mu}_o} T_{\hat{\mu}_1} \cdots T_{\hat{\mu}_{N-1}})(J_o) \leq (T_{\hat{\mu}_o} T_{\hat{\mu}_1} \cdots T_{\hat{\mu}_{N-2}})(TJ_o + \epsilon/N)$$

$$\leq (T_{\hat{\mu}_o} T_{\hat{\mu}_1} \cdots T_{\hat{\mu}_{N-3}})(T^2 J_o + \epsilon/N + \alpha\epsilon/N)$$

$$\leq T^N J_o + \epsilon,$$

where the last inequality is obtained by repeating the process used to obtain the first two inequalities. It follows that $J_N^* \leq T^N J_o$.

If (SM) is summable above, then $J_{K,\pi}(x) < \infty$ for every $x \epsilon S, \pi \epsilon \Pi$, $K=1,\ldots,N$. By (3.6), $T^K J_o(x) < \infty$ for every $x \epsilon S$, $K=1,\ldots,N$. Use Theorem 2.5 to choose nonrandomized policies $\pi^i = (\mu_o^i, \ldots, \mu_{N-1}^i) \epsilon \Pi$ such that for every $x \epsilon S$,

$$T_{\mu_k^i}(T^{N-k-1}J_o)(x) < \infty, \quad k=0,\ldots,N-1, \quad i=1,2,\ldots,$$

and

$$T_{\mu_k^i}(T^{N-k-1}J_o) \downarrow T^{N-k}J_o, \quad k=0,\ldots,N-1,$$

as $i \to \infty$. Then

$$J_N^* \leq \inf_{(i_o,\ldots,i_{N-1})}(T_{\mu_o^{i_o}} \cdots T_{\mu_{N-1}^{i_{N-1}}})(J_o)$$

$$= \inf_{i_o} \cdots \inf_{i_{N-1}} (T_{\mu_o^{i_o}} \cdots T_{\mu_{N-1}^{i_{N-1}}})(J_o)$$

$$= \inf_{i_o} \cdots \inf_{i_{N-2}} (T_{\mu_o^{i_o}} \cdots T_{\mu_{N-2}^{i_{N-2}}})[\inf_{i_{N-1}}(T_{\mu_{N-1}^{i_{N-1}}} J_o)] \text{ by Lemma 3.4}$$

$$= \inf_{i_0} \ldots \inf_{i_{N-2}} (T_{\mu_0^{i_0}} \ldots T_{\mu_{N-2}^{i_{N-2}}})(TJ_0)$$

$$= T^N J_0,$$

where the last equality is obtained by repeating the process used to obtain the previous equality. QED

Corollary 3.2.1 $(F^+)(F^-)$

The function J_N^* is lower semianalytic.

Corollary 3.2.2

(F^+) For each $\epsilon > 0$, there exists a nonrandomized Markov ϵ-optimal policy.

(F^-) For each $\epsilon > 0$, there exists a nonrandomized semi-Markov ϵ-optimal policy and a (randomized) Markov ϵ-optimal policy.

Proof:

If (SM) is summable below, then the policy $(\hat{\mu}_0, \ldots, \hat{\mu}_{N-1})$ constructed in the proof of Theorem 3.2 is ϵ-optimal, nonrandomized and Markov.

Assume (SM) is summable above. We show first the existence of an ϵ-optimal nonrandomized semi-Markov policy. Let $\pi^i = (\mu_0^i, \ldots, \mu_{N-1}^i)$ be as in the proof of Theorem 3.2. Then

$$J_N^* = T^N J_0 = \inf_{(i_0, \ldots, i_{N-1})} (T_{\mu_0^{i_0}} \ldots T_{\mu_{N-1}^{i_{N-1}}})(J_0)$$

$$= \inf_{(i_0, \ldots, i_{N-1})} J_{N, \pi^{(i_0, \ldots, i_{N-1})}},$$

where $\pi^{(i_0, \ldots, i_{N-1})} = (\mu_0^{i_0}, \ldots, \mu_{N-1}^{i_{N-1}})$. Choose $\epsilon > 0$ and define

$$\theta(x) = J^*(x) + \epsilon \text{ if } J^*(x) > -\infty;$$

$$= -1/\epsilon \text{ if } J^*(x) = -\infty.$$

Order linearly the set $\{\pi^{(i_0, \ldots, i_{N-1})}: \ i_0, \ldots, i_{N-1} \text{ are positive integers}\}$

and define $\pi(x)$ to be the first $\pi^{(i_0,\ldots i_{N-1})}$ such that

$$J_{N,\pi^{(i_0,\ldots,i_{N-1})}}(x) \leq \delta(x).$$

Let the components of $\pi(x)$ be $(\mu_0(du_0|x),\mu_1(du_1|x,x_1),\ldots,\mu_{N-1}(du_{N-1}|x,x_{N-1}))$. The sets $\{x: \pi(x)=\pi^{(i_0,\ldots,i_{N-1})}\}$ are universally measurable for each $(i_0,\ldots,i_{N-1})$, so $(\mu_0(du_0|x),\mu_1(du_1|x,x_1),\ldots,\mu_{N-1}(du_{N-1}|x,x_{N-1}))$ is an ϵ-optimal nonrandomized semi-Markov policy.

We now show the existence of an ϵ-optimal (randomized) Markov policy. By Lemma 3.1, there exist $\mu_{N-k}\in U(C|S)$ such that for $k=1,\ldots,N$

$$T_{\mu_{N-k}}(T^{k-1}J_0) \leq (T^kJ_0) + \epsilon/N.$$

Proceed as in (3.7). QED

If (SM) is summable above and $\epsilon>0$, it may not be possible to find an ϵ-optimal nonrandomized Markov policy, as the following example demonstrates.

<u>Example 3.1</u> Let $S=\{0,1,2,\ldots\}$, $C=\{1,2,\ldots\}$, $W=\{w_1,w_2\}$, $\Gamma=SC$, $N=2$ and define

$$\begin{aligned} g(x,u) &= -u \text{ if } x=1,\\ &= 0 \text{ if } x\neq 1, \end{aligned}$$

$$\begin{aligned} f(x,u,w) &= 0 \text{ if } x=0 \text{ or } x=1 \text{ or } w=w_1,\\ &= 1 \text{ if } x\neq 0,\ x\neq 1 \text{ and } w=w_2, \end{aligned}$$

$$\begin{aligned} p(\{w_1\}|x,u) &= 1 - 1/x \text{ if } x\neq 0,\ x\neq 1,\\ p(\{w_2\}|x,u) &= 1/x \quad \text{ if } x\neq 0,\ x\neq 1. \end{aligned}$$

Let $\pi=(\mu_0,\mu_1)$ be a nonrandomized Markov policy. If the initial state x_0 is neither zero nor one, then regardless of the policy employed, $x_1=0$ with probability $1-(1/x_0)$, and $x_1=1$ with probability $1/x_0$. Once the system reaches

zero, it remains there at no further cost. If the system reaches one, it moves to $x_2=0$ at a cost of $-\mu_1(1)$. Thus

$$J_{N,\pi}(x_0) = -\mu_1(1)/x_0 \text{ for } x_0 \neq 0,\ x_0 \neq 1,$$

and

$$J_N^*(x_0) = -\infty \text{ for } x_0 \neq 0,\ x_0 \neq 1.$$

For any $\epsilon>0$, π cannot be ϵ-optimal.

In Example 3.1 it is possible to find a sequence of nonrandomized Markov policies $\{\pi_n\}$ such that $J_{N,\pi_n} \downarrow J_N^*$. This example motivates the idea of policies exhibiting $\{\epsilon_n\}$ dominated convergence to optimality (Definition 3.6) and the following corollary.

<u>Corollary 3.2.3</u> (F^-)

Let $\{\epsilon_n\}$ be a sequence of positive numbers with $\epsilon_n \downarrow 0$. There exists a sequence of nonrandomized Markov policies $\{\pi_n\}$ exhibiting $\{\epsilon_n\}$ dominated convergence to optimality.

Proof:

For N=1, by Theorem 2.5 there exists a sequence of nonrandomized Markov policies $\pi_n=(\mu_0^n)$ for which

$$\begin{aligned}(T_{\mu_0^n}J_0)(x) &\leq (TJ_0)(x) + \epsilon_n \text{ if } (TJ_0)(x)>-\infty;\\ &\leq -1/\epsilon_n \text{ if } (TJ_0)(x)=-\infty.\end{aligned}$$

We may assume without loss of generality that

$$(T_{\mu_0^n}J_0)(x) \downarrow -\infty \text{ if } (TJ_0)(x)=-\infty.$$

Therefore $\{\pi_n\}$ exhibits $\{\epsilon_n\}$ dominated convergence to optimality.

Suppose the result holds for N-1. Let $\pi_n=(\mu_1^n,\dots,\mu_{N-1}^n)$ be a sequence of (N-1)-stage nonrandomized Markov policies exhibiting $\{\epsilon_n/2\alpha\}$ dominated convergence to optimality, i.e.

$$\lim_{n\to\infty} J_{N-1,\pi_n} = J^*_{N-1},$$

$$(3.8)\qquad J_{N-1,\pi_n}(x) \le J_{N-1}(x) + \epsilon_n/2\alpha \text{ if } J^*_{N-1}(x)>-\infty,$$

and

$$(3.9)\qquad J_{N-1,\pi_n}(x) \le J_{N-1,\pi_{n-1}}(x) + \epsilon_n/2\alpha \text{ if } J^*_{N-1}(x)=-\infty.$$

We assume without loss of generality that $\sum_{n=1}^{\infty}\epsilon_n<\infty$. By Theorem 2.5, there exists a sequence $\{\mu^n\}$ of universally measurable functions from S to C whose graphs lie in Γ such that

$$(3.10)\qquad \begin{aligned}(T_{\mu^n}J^*_{N-1})(x) &\le J^*_N(x) + \epsilon_n/2 \text{ if } J^*_N(x)>-\infty;\\ &\le -2/\epsilon_n \text{ if } J^*_N(x)=-\infty.\end{aligned}$$

We may assume without loss of generality that

$$(3.11)\qquad T_{\mu^n}J^*_{N-1} \le T_{\mu^{n-1}}J^*_{N-1},\ n=2,3,\dots.$$

By Theorem 2.4, the set

$$\begin{aligned}A(J^*_{N-1}) &= \{(x,u)\in\Gamma:\ t(J^*_{N-1}=-\infty|x,u)>0\}\\ &= \{(x,u)\in\Gamma:\ \int -\chi_{\{J^*_{N-1}=-\infty\}}(x')t(dx'|x,u)<0\}\end{aligned}$$

is analytic in SC, and von Neumann's Lemma (Theorem A.6) implies the existence of a universally measurable $\mu:\mathrm{proj}_S A(J^*_{N-1})\to C$ whose graph lies in $A(J^*_{N-1})$. Define

$$\begin{aligned}\hat{\mu}^n(x) &= \mu(x) \text{ if } x\in\mathrm{proj}_S A(J^*_{N-1});\\ &= \mu^n(x) \text{ otherwise.}\end{aligned}$$

Then $\hat{\pi}_n=(\hat{\mu}^n,\pi_n)$ is an N-stage nonrandomized Markov policy which we will show exhibits $\{\epsilon_n\}$ dominated convergence to optimality.

For $x\epsilon\mathrm{proj}_S A(J^*_{N-1})$, we have

$$\begin{aligned}\limsup_{n\to\infty} J_{N,\hat{\pi}_n}(x) &= \limsup_{n\to\infty}(T_\mu J_{N-1,\pi_n})(x)\\ &= (T_\mu J^*_{N-1})(x) \text{ by Lemma 3.5}\\ &= -\infty \text{ by choice of } \mu.\end{aligned}$$

For $x\notin\mathrm{proj}_S A(J^*_{N-1})$, we have for every $u\epsilon\Gamma_x$

$$t(J^*_{N-1}=-\infty|x,u) = 0,$$

so by (3.8),

$$\begin{aligned}J_{N,\hat{\pi}_n}(x) &= (T_{\mu^n}J_{N-1,\pi_n})(x)\\ &\leq (T_{\mu^n}J^*_{N-1})(x) + \epsilon_n/2,\end{aligned} \tag{3.12}$$

and

$$\begin{aligned}\limsup_{n\to\infty} J_{N,\hat{\pi}_n}(x) &\leq \limsup_{n\to\infty}(T_{\mu^n}J^*_{N-1})(x)\\ &\leq J^*_N(x)\end{aligned}$$

by (3.10). It follows that

$$\lim_{n\to\infty} J_{N,\hat{\pi}_n} = J^*_N. \tag{3.13}$$

Suppose for fixed $x\epsilon S$, we have $J^*_N(x)>-\infty$. Then $x\notin\mathrm{proj}_S A(J^*_{N-1})$ and we have from (3.10) and (3.12),

$$\begin{aligned}J_{N,\hat{\pi}_n}(x) &\leq (T_{\mu^n}J^*_{N-1})(x) + \epsilon_n/2\\ &\leq J^*_N(x) + \epsilon_n \text{ if } J^*_N(x)>-\infty.\end{aligned} \tag{3.14}$$

Suppose now that $J_N^*(x)=-\infty$. If $x\notin \mathrm{proj}_S A(J_{N-1}^*)$, then (3.11) and (3.12) imply for $n\geq 2$,

$$\begin{aligned} J_{N,\hat{\pi}_n}(x) &\leq (T_{\mu^n}J_{N-1}^*)(x) + \epsilon_n/2 \\ &\leq (T_{\mu^{n-1}}J_{N-1}^*)(x) + \epsilon_n/2 \\ &\leq (T_{\mu^{n-1}}J_{N-1,\pi_{n-1}})(x) + \epsilon_n/2 \\ &\leq J_{N,\hat{\pi}_{n-1}}(x) + \epsilon_n/2, \end{aligned}$$

while if $x\varepsilon \mathrm{proj}_S A(J_{N-1}^*)$, we have from (3.8) and (3.9),

$$\begin{aligned} J_{N,\hat{\pi}_n}(x) &= (T_{\mu}J_{N-1,\pi_n})(x) \\ &\leq (T_{\mu}J_{N-1,\pi_{n-1}})(x) + \epsilon_n/2 \\ &= J_{N,\hat{\pi}_{n-1}}(x) + \epsilon_n/2. \end{aligned}$$

In either case,

$$J_{N,\hat{\pi}_n} \leq J_{N,\hat{\pi}_{n-1}} + \epsilon_n. \tag{3.15}$$

From (3.13), (3.14) and (3.15) we see that $\{\hat{\pi}_n\}$ exhibits $\{\epsilon_n\}$ dominated convergence to optimality. QED

We conclude this section with a technical result needed for the development in Chapter 6.

Corollary 3.2.4 $(F^+)(F^-)$

For every $p\varepsilon P(S)$,

$$\int J_N^*(x)p(dx) = \inf_{\pi\varepsilon\Pi} \int J_{N,\pi}(x)p(dx).$$

Proof:

For $p \in P(S)$,

$$\int J_N^*(x)p(dx) \leq \int J_{N,\pi}(x)p(dx)$$

for every $\pi \in \Pi$, which implies

$$\int J_N^*(x)p(dx) \leq \inf_{\pi \in \Pi} \int J_{N,\pi}(x)p(dx).$$

Choose $\epsilon > 0$ and let $\tilde{\pi} \in \Pi$ be ϵ-optimal. Then

$$\begin{aligned}\int J_{\tilde{\pi},N}(x)p(dx) &\leq \int J_N^*(x)p(dx) + \epsilon \text{ if } p\{x: J_N^*(x)=-\infty\}=0;\\ &\leq -p\{x: J_N^*(x)=-\infty\}/\epsilon \\ &\quad + \int_{\{J_N^*>-\infty\}} J_N^*(x)p(dx) + \epsilon \text{ if } p\{x: J_N^*(x)=-\infty\}>0;\end{aligned}$$

and the reverse inequality follows. QED

<u>Section 3. Existence of Optimal policies</u>

<u>Definition</u> 3.[illegible] Let $\pi=(\mu_0,\ldots,\mu_{N-1})$ be a Markov policy and $\pi^k=(\mu_{N-k},\ldots,\mu_{N-1})$, $k=1,\ldots,N$. The policy π is <u>uniformly N-stage optimal</u> if

$$J_{k,\pi^k} = J_k^*, \quad k=1,\ldots,N.$$

<u>Lemma 3.6</u> $(F^+)(F^-)$

The policy $\pi=(\mu_0,\ldots,\mu_{N-1}) \in \Pi$ is uniformly N-stage optimal if and only if

$$T_{\mu_{N-k}}(T^{k-1}J_0) = T^kJ_0, \quad k=1,\ldots,N.$$

Proof:

If $\pi=(\mu_0,\ldots,\mu_{N-1})$ is uniformly N-stage optimal, then

$$T^kJ_0 = J_k^* = J_{k,\pi^k} = T_{\mu_{N-k}} J_{k-1,\pi^{k-1}} = T_{\mu_{N-k}} J_{k-1}^* = T_{\mu_{N-k}}(T^{k-1}J_0), \quad k=1,\ldots,N.$$

If $T_{\mu_{N-k}}(T^{k-1}J_o) = T^kJ_o$, $k=1,\ldots,N$, then

$$\begin{aligned} J_k^* = T^kJ_o &= T_{\mu_{N-k}}(T^{k-1}J_o) \\ &= T_{\mu_{N-k}}T_{\mu_{N-k+1}}(T^{k-2}J_o) \\ &= (T_{\mu_{N-k}}\cdots T_{\mu_{N-1}})(J_o) \\ &= J_{k,\pi^k},\ k=1,\ldots,N, \end{aligned}$$

where the next to last equality is obtained by continuing the process used to obtain the previous equalities. QED

Theorem 3.3 $(F^+)(F^-)$

If the infimum in

$$\inf_{u\epsilon\Gamma_x}\{g(x,u) + \alpha\int J_k^*(x')t(dx'|x,u)\},\ k=0,\ldots,N-1,$$

is achieved for each $x\epsilon S$, then a uniformly N-stage optimal (and hence optimal) nonrandomized Markov policy exists. This policy is generated by the dynamic programming algorithm, i.e. by measurably selecting for each x a control u which achieves the above infimum.

Proof:

Let $\pi=(\mu_o,\ldots,\mu_{N-1})$, where $\mu_{N-k-1}:S\to C$ achieves the infimum above and satisfies $\mu_{N-k-1}(x)\epsilon\Gamma_x$, for every $x\epsilon S$, $k=0,\ldots,N-1$ (Theorem 2.5). Apply Lemma 3.6. QED

We now make continuity and compactness assumptions on (SM) which guarantee the hypothesis of Theorem 3.3 is satisfied and consequently a uniformly N-stage optimal nonrandomized Markov policy exists. It follows as a by-product of these assumptions and the selection Theorem 2.7 that this policy

can be chosen to be Borel measurable.

(3.16) C is compact.

(3.17) $\Gamma = \bigcup_{j=1}^{\infty} \Gamma^j$, where $\Gamma^1 \subset \Gamma^2 \subset \ldots$, each Γ^j is a closed subset of SC, and

$$\lim_{j\to\infty} \inf_{(x,u)\varepsilon\,\Gamma^j - \Gamma^{j-1}} g(x,u) = +\infty.^1$$

(3.18) g is lower semicontinuous and bounded below on Γ.

(3.19) $t(dx'|x,u)$ is continuous on Γ.

Theorem 3.4 Let (3.16) - (3.19) hold. If $J:S\to R^*$ is lower semicontinuous and bounded below, then so is TJ and there is a Borel measurable function $\mu:S\to C$ such that $\mu(x)\varepsilon\,\Gamma_x$ and

$$(TJ)(x) = g(x,\mu(x)) + \alpha\int J(x')t(dx'|x,\mu(x))$$

for every $x\varepsilon S$.

Proof:

The function

$$(x,u)\to g(x,u) + \alpha\int J(x')t(dx'|x,u)$$

is lower semicontinuous on Γ (Theorem 2.8 and Lemma 2.5(a)). Define

$$H(x,u) = g(x,u) + \alpha\int J(x')t(dx'|x,u) \text{ if } (x,u)\varepsilon\,\Gamma;$$
$$= +\infty \text{ otherwise.}$$

For $c\varepsilon R$, (3.17) and the lower boundedness of J imply the existence of some k such that

[1] By convention the infimum over the empty set is $+\infty$. Thus we allow the possibility that for some k, $\Gamma^k = \Gamma^{k+1} = \Gamma^{k+2} = \ldots$.

$$\inf_{(x,u)\varepsilon \bigcup_{j=k+1}^{\infty} \Gamma^j} H(x,u) > c.$$

Therefore $\{(x,u): H(x,u)\leq c\}=\{(x,u): g(x,u)+\alpha\int J(x')t(dx'|x,u)\leq c\}$ is a subset of Γ^k and closed in Γ^k, therefore closed in SC. It follows that H is lower semicontinuous on SC.

By Theorem 2.6, $(TJ)(x)=\min_{u\varepsilon C}H(x,u)$ is lower semicontinuous. By Theorem 2.7, there exists a Borel measurable $\mu_1:S\rightarrow C$ such that $(TJ)(x)=H(x,\mu_1(x))$ for every $x\varepsilon S$. By Lemma 2.3, there exists a Borel measurable $\mu_2:S\rightarrow C$ such that $(x,\mu_2(x))\varepsilon\Gamma$ for every $x\varepsilon S$. Let

$$\begin{aligned}\mu(x) &= \mu_1(x) \text{ if } (TJ)(x)<\infty;\\ &= \mu_2(x) \text{ otherwise.}\end{aligned}$$

Then μ has the desired properties. QED

We note that although (3.16) requires the compactness of C, the results just proved all hold for noncompact C if the sets Γ^j, j=1,2,..., are compact. This is true because C can be homeomorphically embedded in a compact Borel space $\hat{C}$ [11, Chapter IX, Corollary 9.2] and the images of Γ^j, j=1,2,..., are compact in $S\hat{C}$.

It is also possible to give simple sufficient conditions on g', f and p in Definition 3.1 to insure that (3.18) and (3.19) hold.

<u>Theorem 3.5</u> Suppose that $p(dw|x,u)$ is continuous. If g' is bounded below and lower semicontinuous, then g is also. If f is continuous, then $t(dx'|x,u)$ is also.

Proof:

If p is continuous and g' is bounded below and lower semicontinuous,

Theorem 2.8 implies that g is also. By Theorem B.1, the stochastic kernel t is continuous if and only if for every $F \in C(X)$, the function

$$(x,u) \rightarrow \int F(x')t(dx'|x,u)$$

is continuous. By definition

$$\int F(x')t(dx'|x,u) = \int F[f(x,u,w)]p(dw|x,u),$$

and the conclusion follows from Lemma 2.7. QED

Note that if W is n-dimensional Euclidean space and the distribution of w is given by a density $d(w;x,u)$ which is jointly continuous in (x,u) for fixed w, then $p(dw|x,u)$ is continuous. To see this let G be an open set in W and $(x_n,u_n) \rightarrow (x,u)$ in SC. Then

$$\lim\inf_k p(G|x_k,u_k) = \lim\inf_k \int_G d(w;x_k,u_k)dw \geq \int_G d(w;x,u)dw = p(G|x,u)$$

by Fatou's Lemma. The continuity of $p(dw|x,u)$ follows from Theorem B.1.

In fact it is not necessary that d be continuous in (x,u) for each w, but only that $(x_n,u_n) \rightarrow (x,u)$ imply $d(w;x_n,u_n) \rightarrow d(w;x,u)$ for Lebesgue almost all w. For example, if W=R, the exponential density

$$\begin{aligned} d(w;x,u) &= e^{-(w-m(x,u))} \text{ if } w \geq m(x,u); \\ &= 0 \text{ if } w < m(x,u), \end{aligned}$$

where $m:SC \rightarrow R$ is continuous, has this property.

CHAPTER 4

THE INFINITE HORIZON MODELS AND THEIR RELATIONSHIPS

Section 1. The stochastic model

The stochastic model (SM) considered in this chapter is the same as given by Definition 3.1, except that the horizon N is $+\infty$. The entire discussion of Chapter 3, Section 1, including Theorem 3.1, applies to the infinite horizon model except that in place of the concepts of summability below and summability above we have the assumptions

(P) $g'(x,u,w) \geq 0$ for every $x \in S$, $u \in C$, $w \in W$;

(N) $g'(x,u,w) \leq 0$ for every $x \in S$, $u \in C$, $w \in W$.

We consider additionally the discounted case

(D) $-b \leq g'(x,u,w) \leq b < \infty$ for every $x \in S$, $u \in C$, $w \in W$; $0 < \alpha < 1$.

Any one of these assumptions guarantees the convergence in R^* of the sum in (3.4) when $k \rightarrow \infty$. All results in Chapters 4 and 5 hold in at least one of the cases (P), (N) or (D) and these letters will appear in the statements of the results to indicate which are applicable.

We will also implicitly assume that under (P) any function $J: S \rightarrow R^*$ actually takes values only in $[0,+\infty]$; under (N), J takes values in $[-\infty,0]$; and under (D), J is bounded. One can check that these properties are preserved by the operators T_μ and T (Definitions 3.7 and 3.8). By making these assumptions we ensure that the integral operates linearly.

When the horizon is infinite, we omit the subscript N in the functions $J_{N,\pi}$ and J_N^*. In terms of the operator T_μ, we have in place of equality in (3.5) for $\pi=(\mu_0,\mu_1,\ldots)\varepsilon\Pi$,

$$J_\pi = \lim_{N\to\infty} (T_{\mu_0}\ldots T_{\mu_{N-1}})(J_0).$$

If $\pi=(\mu,\mu,\ldots)\varepsilon\Pi$ has all components the same, we say π is <u>stationary</u> and write J_μ in place of J_π.

<u>Section 2. The deterministic model</u>

We now describe the deterministic decision model (DM) corresponding to (SM).

<u>Definition 4.1</u> Let $(S,C,\Gamma,W,p,f,\alpha,g',N)$ be a stochastic decision model as described by Definition 3.1, let $N=\infty$, and let t be the transition kernel defined by (3.1) and g the (reduced) one-stage cost function defined by (3.2). The corresponding <u>deterministic model (DM)</u> consists of the following:

$P(S)$: <u>State space</u>.

$P(SC)$: <u>Control space</u>.

$P(\Gamma)=\{q\varepsilon P(SC): q(\Gamma)=1\}$: <u>Constraint set</u>. For $p\varepsilon P(X)$,

$$P(\Gamma)_p=\{q\varepsilon P(\Gamma): \text{ The marginal of } q \text{ on } S \text{ is } p.\}$$

$\bar{f}$: <u>System function</u>. $\bar{f}:P(SC)\to P(S)$ is defined by

$$\bar{f}(q)(\underline{S}) = \int_{SC} t(\underline{S}|x,u)q(d(x,u)),\ \underline{S}\varepsilon\mathcal{B}_S.$$

α,g: <u>Discount factor</u> and <u>one-stage cost function</u>. We treat the cases (P), (N) and (D).

By Corollary B.6.1, $P(S)$, $P(C)$ and $P(SC)$ are Borel spaces. By Theorem B.6, $P(\Gamma)=\{q\varepsilon P(SC): q(\Gamma)\geq 1\}$ is an analytic subset of $P(SC)$. By Corollaries B.3.2 and B.3.3, $\bar{f}$ is Borel measurable. The function $q \rightarrow \int g\, dq$ is lower semianalytic on $P(SC)$ by Corollary 2.4.1.

Definition 4.2 A policy in (DM) is a sequence of mappings $\bar{\pi}=(\bar{\mu}_0,\bar{\mu}_1,\ldots)$ such that for each k, $\bar{\mu}_k: P(S)\rightarrow P(SC)$ and $\bar{\mu}_k(p)\, \varepsilon\, P(\Gamma)_p$ for every $p\varepsilon P(S)$. The set of all policies in (DM) will be denoted by $\bar{\Pi}$. We place no measurability requirements on these mappings. If $\bar{\pi}=(\bar{\mu},\bar{\mu},\ldots)$ has all components the same, $\bar{\pi}$ is said to be stationary.

Definition 4.3 Given $p_0\varepsilon P(S)$ and a policy $\bar{\pi}$ in (DM), the cost function corresponding to $\bar{\pi}$ at p_0 is

$$\bar{J}_{\bar{\pi}}(p_0) = \sum_{k=0}^{\infty} \alpha^k \int g\, dq_k, \tag{4.1}$$

where the q_k's are generated recursively by

$$q_k = \bar{\mu}_k(p_k), \tag{4.2}$$

$$p_{k+1} = \bar{f}(q_k),\quad k=0,1,\ldots. \tag{4.3}$$

If $\bar{\pi}=(\bar{\mu},\bar{\mu},\ldots)$ is stationary, we write $\bar{J}_{\bar{\mu}}$ in place of $\bar{J}_{\bar{\pi}}$. The optimal cost function at p_0 is

$$\bar{J}^*(p_0) = \inf_{\bar{\pi}\varepsilon\bar{\Pi}} \bar{J}_{\bar{\pi}}(p_0).$$

The concepts of ϵ-optimal and optimal policies for (DM) are the same as those given in Definition 3.6 for (SM).

Definition 4.4 A sequence $(p_0,q_0,q_1,\ldots)\, \varepsilon\, P(S)P(SC)P(SC)\ldots$ is admissible in (DM) if $q_0 \varepsilon P(\Gamma)_{p_0}$ and $q_{k+1}\varepsilon P(\Gamma)_{\bar{f}(q_k)}$, $k=0,1,\ldots$. The set of all admissible sequences will be denoted by Δ.

The admissible sequences are just those which can be generated by some policy via (4.2) and (4.3). Except for p_o, the p_k's are not included in the sequence, but these are the marginals of the q_k's on S and so can be recovered from the sequence.

<u>Lemma 4.1</u> The set of admissible sequences Δ in (DM) is an analytic subset of $P(S)P(SC)P(SC)\ldots$.

Proof:

The set Δ is equal to $A_o \cap [\bigcap_{k=0}^{\infty} B_k]$, where

$$A_o = \{(p_o,q_o,q_1,\ldots): q_o \varepsilon P(\Gamma)_{p_o}\},$$
$$B_k = \{(p_o,q_o,q_1,\ldots): q_{k+1} \varepsilon P(\Gamma)_{\bar{f}(q_k)}\}.$$

To show that A_o and B_k, $k=0,1,\ldots$, are analytic, by Theorem A.2 it suffices to show

$$A = \{(p_1,q_1) \varepsilon P(S)P(SC): q_1 \varepsilon P(\Gamma)_{p_1}\}$$

and

$$B = \{(q_o,q_1) \varepsilon P(SC)P(SC): q_1 \varepsilon P(\Gamma)_{\bar{f}(q_o)}\}$$

are analytic.

The set A is the intersection of $P(S)P(\Gamma)$ with the graph of the continuous mapping $q \rightarrow$ [Marginal of q] from $P(SC)$ to $P(S)$, and this is analytic by Theorem A.2 and [26], Chapter 1, Theorem 3.3. The set B is the inverse image of A under the Borel measurable mapping

$$(q_o,q_1) \rightarrow (\bar{f}(q_o),q_1),$$

and so is analytic (Theorem A.3). QED

Section 3. Relations between the models

The deterministic model lends itself to a simpler analysis than does the stochastic model and Chapter 5 is devoted to this. For the analysis to be meaningful, however, it is necessary to establish correspondences between (SM) and (DM) that permit transfer of results from one to the other. This is the goal of this section.

Definition 4.5 Let $\pi=(\mu_0,\mu_1,\ldots)\epsilon\Pi$ be a policy in (SM) and $\bar{\pi}=(\bar{\mu}_0,\bar{\mu}_1,\ldots)$ a policy in (DM). Let p_0 be given. If for each $\underline{S}\epsilon\mathcal{B}_S$, $\underline{C}\epsilon\mathcal{B}_C$,

$$(4.4) \qquad \int_{\underline{S}}\mu_k(\underline{C}|x)p_k(dx) = \bar{\mu}_k(p_k)(\underline{SC}),$$

where the p_k's are generated from p_0 by $\bar{\pi}$ via (4.2) and (4.3), then π and $\bar{\pi}$ correspond at p_0. If π and $\bar{\pi}$ correspond at every $p\epsilon P(S)$, then π and $\bar{\pi}$ correspond.

If π and $\bar{\pi}$ correspond at p_0, then the sequence of measures $(q_0,q_1,\ldots)$ generated from p_0 by π via (3.3) is the same as the sequence generated from p_0 by $\bar{\pi}$ via (4.2) and (4.3). If π and $\bar{\pi}$ correspond, then they generate the same sequence $(q_0,q_1,\ldots)$ for any initial p_0.

Theorem 4.1 (P)(N)(D)

If $\pi\epsilon\Pi$, there is a corresponding $\bar{\pi}\epsilon\bar{\Pi}$. If $\bar{\pi}\epsilon\bar{\Pi}$ and $p_0\epsilon P(S)$ is given, there is a policy $\pi\epsilon\Pi$ corresponding to $\bar{\pi}$ at p_0.

Proof:

In the first case, let $\bar{\pi}=(\bar{\mu}_0,\bar{\mu}_1,\ldots)$, where $\bar{\mu}_k$ is chosen to satisfy (4.4) identically in p_k. In the second case, let $\pi=(\mu_0,\mu_1,\ldots)$, where μ_k is chosen to satisfy (4.4) when the p_k's are generated from p_0 by $\bar{\pi}$ via (4.2) and (4.3). Such μ_k's exist by Theorem C.1. QED

Theorem 4.2 (P)(N)(D)

If $\pi\varepsilon\Pi$ and $\bar{\pi}\varepsilon\bar{\Pi}$ correspond at p, then

$$\int J_{\pi}(x)p(dx) = \bar{J}_{\bar{\pi}}(p).$$

Proof:

Let $(q_0,q_1,\ldots)$ be generated from p by π via (3.3) or from p by $\bar{\pi}$ via (4.2) and (4.3). Then

$$\bar{J}_{\bar{\pi}}(p) = \sum_{k=0}^{\infty}\alpha^k\int g\, dq_k = \int J_{\pi}(x)p(dx),$$

where the monotone or bounded convergence theorem is used to interchange integration and summation for the second equality. QED

Corollary 4.2.1 (P)(N)(D)

If $\pi\varepsilon\Pi$ and $\bar{\pi}\varepsilon\bar{\Pi}$ correspond at p_x, then $J_{\pi}(x) = \bar{J}_{\bar{\pi}}(p_x)$.

Corollary 4.2.2 (P)(N)(D)

For every $x\varepsilon S$, $J^*(x)=\bar{J}^*(p_x)$.

Proof:

This is an immediate consequence of Theorems 3.1 and 4.1 and Corollary 4.2.1. QED

Definition 4.6 Let $\bar{J}:P(S)\longrightarrow R^*$ and $\bar{\mu}:P(S)\longrightarrow P(SC)$ be such that $\bar{\mu}(p)\,\varepsilon\, P(\Gamma)_p$ for every $p\varepsilon P(S)$. The operator $\bar{T}_{\bar{\mu}}$ mapping $\bar{J}$ into $\bar{T}_{\bar{\mu}}\bar{J}:P(S)\longrightarrow R^*$ is defined by

$$(\bar{T}_{\bar{\mu}}\bar{J})(p) = \int g\, d\bar{\mu}(p) + \alpha\bar{J}[\bar{f}(\bar{\mu}(p))]$$

for every $p\varepsilon P(S)$. The operator $\bar{T}$ mapping $\bar{J}$ into $\bar{T}\bar{J}:P(S)\longrightarrow R^*$ is defined by

$$(\bar{T}\bar{J})(p) = \inf_{q\varepsilon P(\Gamma)_p}\{\int g\, dq + \alpha \bar{J}[\bar{f}(q)]\}.$$

We implicitly assume that under (P) any function $\bar{J}:P(S)\rightarrow R^*$ actually takes values only in $[0,+\infty]$; under (N), $\bar{J}$ takes values in $[-\infty,0]$; and under (D), $\bar{J}$ is bounded. One can verify that these properties are preserved by the operators $\bar{T}_{\bar{\mu}}$ and $\bar{T}$.

The next theorem is a direct consequence of the definitions.

<u>Theorem</u> <u>4.3</u> (P)(N)(D)

Let $J:S\rightarrow R^*$ be universally measurable and $\bar{J}(p)=\int J dp$. Suppose $\bar{\mu}:P(S)\rightarrow P(SC)$ is of the form

$$\bar{\mu}(p)(\underline{SC}) = \int_{\underline{S}}\mu(\underline{C}|x)p(dx),\ \underline{S}\varepsilon\mathcal{B}_S,\ \underline{C}\varepsilon\mathcal{B}_C,$$

for some universally measurable stochastic kernel $\mu\varepsilon U(C|S)$.[1] Then $\bar{\mu}(p)\ \varepsilon\ P(\Gamma)_p$ and

$$(\bar{T}_{\bar{\mu}}\bar{J})(p) = \int(T_\mu J)dp$$

for every $p\varepsilon P(S)$.

<u>Theorem</u> <u>4.4</u> (P)(N)(D)

Let $J:S\rightarrow R^*$ be lower semianalytic and $\bar{J}(p)=\int J dp$. Then

$$(\bar{T}\bar{J})(p) = \int(TJ)dp$$

for every $p\varepsilon P(S)$.

Proof:

For $q\varepsilon P(\Gamma)_p$,

[1]The set $U(C|S)$ is defined in Chapter 3, Section 2.

$$\int g\,dq + \alpha\bar{J}[\bar{f}(q)] = \int [g(x,u) + \alpha\int J(x')t(dx'|x,u)]q(d(x,u))$$
$$\geq \int (TJ)(x)p(dx),$$

which implies

$$(\overline{\overline{T}}J)(p) \geq \int (TJ)dp.$$

Given $p\varepsilon P(S)$ and $\epsilon>0$, by Theorem 2.5 there is a universally measurable selector $\mu:S\rightarrow C$ such that $(x,\mu(x))\varepsilon\Gamma$ for every $x\varepsilon S$ and

$$g(x,\mu(x)) + \alpha\int J(x')t(dx'|x,\mu(x))$$
$$\leq (TJ)(x) + \epsilon \text{ if } (TJ)(x)>-\infty;$$
$$\leq -(1+\epsilon^2)/\epsilon\, p\{x:\ (TJ)(x)=-\infty\} \text{ if } (TJ)(x)=-\infty \text{ and } p\{x:\ (TJ)(x)=-\infty\}>0.$$

Let $q\varepsilon P(\Gamma)_p$ be defined by

$$q(\underline{S}\underline{C}) = \int_{\underline{S}} p_{\mu(x)}(\underline{C})p(dx),\quad \underline{S}\varepsilon\mathcal{B}_S,\ \underline{C}\varepsilon\mathcal{B}_C.$$

Then

$$\int g\,dq + \alpha\bar{J}[\bar{f}(q)] \leq \int (TJ)(x)p(dx) + \epsilon \text{ if } p\{x:\ (TJ)(x)=-\infty\}=0;$$
$$\leq -1/\epsilon \text{ if } p\{x:\ (TJ)(x)=-\infty\}>0.$$

Therefore $(\overline{\overline{T}}J)(p)\leq\int(TJ)dp$. QED

Corollary 4.2.2 has shown that J^* and $\bar{J}^*$ are related, but in a rather weak way that involves $\bar{J}^*$ only on $\bar{S}=\{p_x\varepsilon P(S):\ x\varepsilon S\}$. In Theorem 4.6 we strengthen this relationship, but in order to state that theorem we must show a measurability property of $\bar{J}^*$.

<u>Theorem 4.5</u> (P)(N)(D)

The function $\bar{J}^*$ is lower semianalytic on $P(S)$.

Proof:

Define

$$G(p_o,q_o,q_1,\ldots) = \sum_{k=0}^{\infty}\alpha^k\int g\, dq_k \text{ if } (p_o,q_o,\ldots)\in\Delta;$$
$$= +\infty \text{ otherwise.}$$

Then G is lower semianalytic on P(S)P(SC)P(SC)... by the remark following Definition 4.1 and by Lemma 4.1. By Theorem 2.3,

$$\bar{J}^*(p_o) = \inf_{(p_o,q_o,\ldots)\in P(S)P(SC)\ldots} G(p_o,q_o,q_1,\ldots)$$

is lower semianalytic. QED

<u>Corollary 4.5.1</u> (P)(N)(D)

The function J^* is lower semianalytic on S.

Proof:

By Corollary B.7.1, $\theta: x \longrightarrow p_x$ is continuous from S onto $\bar{S}$. By Corollary 4.2.2, $J^*(x)=\bar{J}^*(\theta(x))$, and so $\{x: J^*(x)<c\}=\theta^{-1}\{p: \bar{J}^*(p)<c\}$ is analytic (Theorem A.3). QED

<u>Lemma 4.2</u> (P)(N)(D)

Given $p\in P(S)$ and $\epsilon>0$, there exists a policy $\bar{\pi}$ in (DM) such that

(P) (D) $\qquad \bar{J}_{\bar{\pi}}(p) \le \int J^*(x)p(dx) + \epsilon;$

(N) $\qquad \bar{J}_{\bar{\pi}}(p) \le \int J^*(x)p(dx) + \epsilon$ if $p\{x: J^*(x)=-\infty\}=0;$

$\qquad \le -1/\epsilon$ if $p\{x: J^*(x)=-\infty\}>0.$

Proof:

Let $\bar{p}\in P(S)$ and $\epsilon>0$ be given. Let G be as defined in the proof of Theorem 4.5.

Under (P) and (D), by Theorem 2.5 there exists a universally measurable selector $\varphi: P(S) \to P(SC)P(SC)\ldots$ such that $(p,\varphi(p)) \in \Delta$ and $G(p,\varphi(p)) \leq \bar{J}^*(p)+\epsilon$ for every p. Letting $\theta: x \to p_x$ and $s(x)=\varphi(\theta(x))$, we have $(p_x, s(x)) \in \Delta$ and $G(p_x, s(x)) \leq J^*(x)+\epsilon$ for every $x \in S$. By Corollary 2.2.1, s is universally measurable.

Under (N), select s universally measurable so that for every $x \in S$, we have $(p_x, s(x)) \in \Delta$ and

$$
\begin{aligned}
G(p_x, s(x)) &\leq J^*(x) + \epsilon \text{ if } J^*(x) > -\infty; \\
&\leq -(1+\epsilon^2)/\epsilon\, \bar{p}\{x: J^*(x)=-\infty\} \text{ if } J^*(x)=-\infty \text{ and } \bar{p}\{x: J^*(x)=-\infty\}>0.
\end{aligned}
$$

Denote $s(x_o)=(q_o(\cdot|x_o), q_1(\cdot|x_o),\ldots)$. Each $q_k(\cdot|x_o)$ is a universally measurable stochastic kernel on SC given S (Theorem C.3) satisfying

$$q_{k+1}(\cdot|x_o) \in P(\Gamma)_{\bar{f}[q_k(\cdot|x_o)]}, \quad k=0,1,\ldots,$$

for each $x_o \in S$. Define $\bar{q}_k \in P(SC)$ by

$$\bar{q}_k(B) = \int q_k(B|x_o)\bar{p}(dx_o), \quad B \in \mathcal{B}_{SC}$$

Then $\bar{q}_k \in P(\Gamma)$, $k=0,1,\ldots$. We show that $(\bar{p},\bar{q}_o,\bar{q}_1,\ldots) \in \Delta$. For $\underline{S} \in \mathcal{B}_S$, $k=1,2,\ldots$

$$
\begin{aligned}
\bar{q}_k(\underline{S}C) &= \int_S q_k(\underline{S}C|x_o)\bar{p}(dx_o) \\
&= \int_S \int_{SC} t(\underline{S}|x,u) q_{k-1}(d(x,u)|x_o)\bar{p}(dx_o) \\
&= \int_{SC} t(\underline{S}|x,u)\bar{q}_{k-1}(d(x,u)),
\end{aligned}
$$

which implies $\bar{q}_k \in P(\Gamma)_{\bar{f}(\bar{q}_{k-1})}$. For $k=0$,

$$\bar{q}_o(\underline{S}C) = \int_S q_o(\underline{S}C|x_o)\bar{p}(dx_o) = \int_S \chi_{\underline{S}}(x_o)\bar{p}(dx_o) = \bar{p}(\underline{S}).$$

Let $\bar{\pi}$ be any policy in (DM) which generates the admissible sequence $(\bar{p},\bar{q}_o,\bar{q}_1,\ldots)$. Then under (P) and (D)

$$\bar{J}_{\bar{\pi}}(\bar{p}) = G(\bar{p},\bar{q}_o,\bar{q}_1,\ldots) = \int_S [\sum_{k=0}^{\infty} \alpha^k \int_{SC} g(x,u) q_k(d(x,u)|x_o)] \bar{p}(dx_o)$$

$$\leq \int_S G(p_{x_o}, s(x_o))\bar{p}(dx_o) \leq \int J^*(x_o)\bar{p}(dx_o) + \epsilon,$$

while under (N)

$$\bar{J}_{\bar{\pi}}(\bar{p}) \leq \int_S G(p_{x_o}, s(x_o))\bar{p}(dx_o) \leq \int J^*(x_o)\bar{p}(dx_o) + \epsilon \text{ if } \bar{p}\{x: J^*(x)=-\infty\}=0;$$

$$\leq -1/\epsilon \text{ if } \bar{p}\{x: J^*(x)=-\infty\}>0. \quad \text{QED}$$

<u>Theorem</u> <u>4.6</u> (P)(N)(D)

For every $p\epsilon P(S)$, $\bar{J}^*(p)=\int J^*(x)p(dx)$.

Proof:

Lemma 4.2 shows that $\bar{J}^*(p)\leq\int J^*(x)p(dx)$. For the reverse inequality, let p be in $P(S)$ and $\bar{\pi}\,\epsilon\,\bar{\Pi}$. Let π be a policy in Π corresponding to $\bar{\pi}$ at p. Then by Theorem 4.2,

$$\bar{J}_{\bar{\pi}}(p) = \int J_\pi(x)p(dx) \geq \int J^*(x)p(dx),$$

and infimizing over $\bar{\pi}\,\epsilon\,\bar{\Pi}$ the theorem follows. QED

<u>Corollary</u> <u>4.6.1</u> (P)(N)(D)

Suppose $\pi\epsilon\Pi$ and $\bar{\pi}\,\epsilon\,\bar{\Pi}$ are corresponding policies for (SM) and (DM). Then π is optimal if and only if $\bar{\pi}$ is optimal.

Proof:

If π is optimal, then for $p\epsilon P(S)$,

$$\bar{J}_{\bar{\pi}}(p) = \int J_\pi(x)p(dx) = \int J^*(x)p(dx) = \bar{J}^*(p).$$

If $\bar{\pi}$ is optimal, then for $x \in S$,

$$J_{\bar{\pi}}(x) = \bar{J}_{\bar{\pi}}(p_x) = \bar{J}^*(p_x) = J^*(x). \qquad \text{QED}$$

We conclude with a technical result needed for Chapter 6.

<u>Corollary 4.6.2</u> (P)(N)(D)

For every $p \in P(S)$,

$$\int J^*(x)p(dx) = \inf_{\pi\in\Pi}\int J_\pi(x)p(dx).$$

Proof:

By Theorems 4.1 and 4.2,

$$\inf_{\pi\in\Pi}\int J_\pi(x)p(dx) = \bar{J}^*(p).$$

Apply Theorem 4.6. QED

CHAPTER 5

MAIN RESULTS - INFINITE HORIZON

Section 1. Introduction

In this chapter we treat cases (P),(N) and (D) of Chapter 4. Due to the infinite horizon, these conditions have been imposed on the one-stage cost function and/or the discount factor to insure that the cost function is well-defined in R^*. Because of these conditions, it was possible to establish strong connections between (SM) and (DM), in particular, Theorems 4.1 - 4.4 and 4.6. These connections will now be exploited.

Since we refer frequently to Bertsekas [4] in this chapter, we begin by pointing out how (DM) satisfies assumptions made in that paper. This can be done because there are no measurability restrictions on the policies in (DM).

Bertsekas considers a mapping H. In our case the arguments of H are $p\varepsilon P(S)$, $q\varepsilon P(SC)$ and $\bar{J}:P(S)\rightarrow R^*$. We define

$$H(p,q,\bar{J}) = \int g\, dq + \alpha\bar{J}[\bar{f}(q)],$$

and then $\bar{T}_{\bar{\mu}}$ and $\bar{T}$ as given in our Definition 4.6 correspond to Bertsekas' T_μ and T.

It is easily verified that for H thus defined, Bertsekas' monotonicity assumption holds. Furthermore, taking $\bar{J}$ to be identically zero in his definitions of the value function corresponding to a policy and the optimal value function, we obtain our definitions of the cost function corresponding to a policy and the optimal cost function given in Chapter 4. Our case (D) corresponds to his contraction assumption, our case (P) to his uniform

increase assumption, and our case (N) to his uniform decrease assumption. His additional assumptions I.1, I.2, D.1 and D.2 are satisfied in the appropriate cases in our model.

Results similar to those of [4] and adequate for our purposes can be found in Bertsekas [3], Chapters 6 and 7. We will give both references whenever possible.

Section 2. The optimality equations and characterizations of optimal policies

Theorem 5.1 (Optimality equations) (P)(N)(D)

We have

$$\bar{J}^* = \bar{T}\bar{J}^*, \quad J^* = TJ^*.$$

Proof:

This holds for (DM) by [4], Propositions 1, 5 and 6 or by [3], Chapter 6, Proposition 2, and Chapter 7, Proposition 1. For (SM) we have for each $x \in S$,

$$J^*(x) = \bar{J}^*(p_x) = (\bar{T}\bar{J}^*)(p_x) = (TJ^*)(x)$$

by Theorems 4.4, 4.6 and Corollary 4.5.1. QED

Theorem 5.2 (P)(N)(D)

If $\bar{\pi}=(\bar{\mu},\bar{\mu},\ldots)$ is a stationary policy in (DM), then $\bar{J}_{\bar{\mu}}=\bar{T}_{\bar{\mu}}\bar{J}_{\bar{\mu}}$. If $\pi=(\mu,\mu,\ldots)$ is a stationary policy in (SM), then $J_\mu=T_\mu J_\mu$.

Proof:

This follows for (DM) by [4], Proposition 1 and Corollaries 5.1 and 6.2 or by [3], Chapter 6, Corollary 2.1, and Chapter 7, Corollary 1.1. Let $\pi=(\mu,\mu,\ldots)$ be a stationary policy in (SM) and $\bar{\pi}=(\bar{\mu},\bar{\mu},\ldots)$ correspond to π.

Then for each $x \in S$,

$$J_\mu(x) = \bar{J}_{\bar{\mu}}(p_x) = (\bar{T}_{\bar{\mu}}\bar{J}_{\bar{\mu}})(p_x) = (T_\mu J_\mu)(x)$$

by Theorems 4.2 and 4.3. QED

<u>Theorem 5.3</u>

(P) If $\bar{J}:P(S)\to[0,+\infty]$ and $\bar{J}\geq\bar{T}\bar{J}$, then $\bar{J}\geq\bar{J}^*$.

If $J:S\to[0,+\infty]$ is lower semianalytic and $J\geq TJ$, then $J\geq J^*$.

(N) If $\bar{J}:P(S)\to[-\infty,0]$ and $\bar{J}\leq\bar{T}\bar{J}$, then $\bar{J}\leq\bar{J}^*$.

If $J:S\to[-\infty,0]$ is lower semianalytic and $J\leq TJ$, then $J\leq J^*$.

(D) If $\bar{J}:P(S)\to[-c,+c]$, $c<\infty$, and $\bar{J}=\bar{T}\bar{J}$, then $\bar{J}=\bar{J}^*$.

If $J:S\to[-c,+c]$, $c<\infty$, is lower semianalytic and $J=TJ$, then $J=J^*$.

Proof:

The proof for (DM) is adapted from [3], Chapter 6, Proposition 9. Under (P) given $\bar{p}_0\in P(S)$ and $\epsilon>0$, choose a sequence $\epsilon_k>0$ such that $\sum_{k=0}^{\infty}\alpha^k\epsilon_k<\epsilon$. Choose $(\bar{q}_0,\bar{q}_1,\ldots)$ such that $(\bar{p}_0,\bar{q}_0,\bar{q}_1,\ldots)\in\Delta$ and

$$\int g\, d\bar{q}_k + \alpha\bar{J}[\bar{f}(\bar{q}_k)] \leq (\bar{T}\bar{J})(\bar{p}_k) + \epsilon_k,$$

where $\bar{p}_k=\bar{f}(\bar{q}_{k-1})$, $k=1,2,\ldots$. Then

$$\bar{J}^*(\bar{p}_0) = \inf_{(p_0,q_0,\ldots)\in\Delta,\, p_0=\bar{p}_0} \sum_{k=0}^{\infty}\alpha^k\int g\, dq_k$$

$$\leq \liminf_{N\to\infty}[\alpha^N\bar{J}(\bar{p}_N)+\sum_{k=0}^{N-1}\alpha^k\int g\, d\bar{q}_k].$$

By choice of the q_k's,

$$\alpha^N\bar{J}[\bar{p}_N] + \sum_{k=0}^{N-1}\alpha^k\int g\, d\bar{q}_k \leq \alpha^{N-1}(\bar{T}\bar{J})(\bar{p}_{N-1}) + \sum_{k=0}^{N-2}\alpha^k\int g\, d\bar{q}_k + \alpha^{N-1}\epsilon_{N-1}$$

$$\leq \alpha^{N-1}\bar{J}(\bar{p}_{N-1}) + \sum_{k=0}^{N-2}\alpha^k\int g\, d\bar{q}_k + \alpha^{N-1}\epsilon_{N-1}$$

$$\leq \bar{J}(\bar{p}_o) + \sum_{k=0}^{N-1} \alpha^k \epsilon_k$$

$$\leq \bar{J}(\bar{p}_o) + \epsilon,$$

where the next to last inequality is obtained by repeating the process used to obtain the previous inequalities. Therefore $\bar{J}^*(\bar{p}_o) \leq \bar{J}(\bar{p}_o) + \epsilon$ and the result follows.

Under (N), for $\bar{p}_o \varepsilon P(S)$,

$$\bar{J}^*(\bar{p}_o) = \inf_{(p_o,q_o,\ldots)\varepsilon\Delta, p_o=\bar{p}_o} \sum_{k=0}^{\infty} \alpha^k \int g \, dq_k$$

$$\geq \inf_{(p_o,q_o,\ldots)\varepsilon\Delta, p_o=\bar{p}_o} \limsup_{N\to\infty} [\alpha^N \bar{J}(p_N) + \sum_{k=0}^{N-1} \alpha^k \int g \, dq_k]$$

$$\geq \limsup_{N\to\infty} \inf_{(p_o,q_o,\ldots)\varepsilon\Delta, p_o=\bar{p}_o} [\alpha^N \bar{J}(p_N) + \sum_{k=0}^{N-1} \alpha^k \int g \, dq_k],$$

where $p_N = \bar{f}(q_{N-1})$, $N=1,2,\ldots$. Now

$$\inf_{(p_o,q_o,\ldots)\varepsilon\Delta, p_o=\bar{p}_o} [\alpha^N \bar{J}(p_N) + \sum_{k=0}^{N-1} \alpha^k \int g \, dq_k]$$

$$= \inf_{(p_o,q_o,\ldots)\varepsilon\Delta, p_o=\bar{p}_o} [\sum_{k=0}^{N-2} \alpha^k \int g \, dq_k$$

$$+ \alpha^{N-1} \inf_{q_{N-1}\varepsilon P(\Gamma)_{p_{N-1}}} \{\int g dq_{N-1} + \alpha \bar{J}[\bar{f}(q_{N-1})]\}]$$

$$= \inf_{(p_o,q_o,\ldots)\varepsilon\Delta, p_o=\bar{p}_o} [\alpha^{N-1} (\bar{T}\bar{J})(p_{N-1}) + \sum_{k=0}^{N-2} \alpha^k \int g \, dq_k]$$

$$\geq \inf_{(p_o,q_o,\ldots)\varepsilon\Delta, p_o=\bar{p}_o} [\alpha^{N-1} \bar{J}(p_{N-1}) + \sum_{k=0}^{N-2} \alpha^k \int g \, dq_k]$$

$$\geq \bar{J}(\bar{p}_o),$$

where the last inequality is obtained by repeating the process used to obtain the previous inequalities, and the result follows.

Under (D), since $0<\alpha<1$ and $\bar{J}$ is bounded,

$$\lim_{N\to\infty}\alpha^N\bar{J}(p_N) = 0,$$

and both the previous arguments can be used.

We now establish the (SM) part of the theorem for (P). Cases (N) and (D) are shown in the same manner.

Under (P) define $\bar{J}(p)=\int Jdp$. Then

$$\bar{J}(p) = \int Jdp \geq \int (TJ)dp = (\bar{T}\bar{J})(p)$$

by Theorem 4.4. By the result for (DM), $\bar{J}\geq\bar{J}^*$. In particular,

$$J(x) = \bar{J}(p_x) \geq \bar{J}^*(p_x) = J^*(x). \qquad \text{QED}$$

Theorem 5.4 Let $\bar{\pi}=(\bar{\mu},\bar{\mu},\ldots)$ and $\pi=(\mu,\mu,\ldots)$ be stationary policies in (DM) and (SM) respectively.

(P) If $\bar{J}:P(S)\to[0,+\infty]$ and $\bar{J}\geq\bar{T}_{\bar{\mu}}\bar{J}$, then $\bar{J}\geq\bar{J}_{\bar{\mu}}$.

If $J:S\to[0,+\infty]$ is universally measurable and $J\geq T_\mu J$, then $J\geq J_\mu$.

(N) If $\bar{J}:P(S)\to[-\infty,0]$ and $\bar{J}\leq\bar{T}_{\bar{\mu}}\bar{J}$, then $\bar{J}\leq\bar{J}_{\bar{\mu}}$.

If $J:S\to[-\infty,0]$ is universally measurable and $J\leq T_\mu J$, then $J\leq J_\mu$.

(D) If $\bar{J}:P(S)\to[-c,+c]$, $c<\infty$, and $\bar{J}=\bar{T}_{\bar{\mu}}\bar{J}$, then $\bar{J}=\bar{J}_{\bar{\mu}}$.

If $J:S\to[-c,+c]$, $c<\infty$, is universally measurable and $J=T_\mu J$, then $J=J_\mu$.

Proof:

The proof for (DM) is a simplification of the one used for Theorem 5.3. The proof for (SM) then follows from Theorem 4.3 and Corollary 4.2.1. QED

Theorem 5.4 implies that under (P), J_μ is the smallest nonnegative universally measurable solution to the functional equation

$$J = T_\mu J.$$

Under (D), J_μ is the only bounded universally measurable solution to this equation. This provides us with a simple necessary and sufficient condition for a stationary policy to be optimal under (P) and (D).

<u>Theorem 5.5</u> (P)(D)

Let $\bar{\pi}=(\bar{\mu},\bar{\mu},\ldots)$ and $\pi=(\mu,\mu,\ldots)$ be stationary policies in (DM) and (SM) respectively. The policy $\bar{\pi}$ is optimal if and only if $\bar{J}^*=\bar{T}_{\bar{\mu}}\bar{J}^*$. The policy π is optimal if and only if $J^*=T_\mu J^*$.

Proof:

The proof for (DM) can be found in [4] or [3], but given the previous theorems, it is quite simple, so we repeat it here.

If $\bar{\pi}$ is optimal, $\bar{J}_{\bar{\mu}}=\bar{J}^*$. By Theorem 5.2, $\bar{J}^*=\bar{T}_{\bar{\mu}}\bar{J}^*$. Conversely, if $\bar{J}^*=\bar{T}_{\bar{\mu}}\bar{J}^*$, then by Theorem 5.4, $\bar{J}^*\geq\bar{J}_{\bar{\mu}}$ and $\bar{\pi}$ is optimal. The proof for (SM) follows from the (SM) parts of the same theorems. QED

<u>Corollary 5.5.1</u> (P)(D)

There is an optimal nonrandomized stationary policy if and only if for each $x\in S$ the infimum in

$$\inf_{u\in\Gamma_x}[g(x,u) + \alpha\int J^*(x')t(dx'|x,u)]$$

is achieved.

Proof:

If the above infimum is achieved for every $x\in\Gamma_x$, then by Theorem 2.5 there is a universally measurable selector $\mu:S\rightarrow C$ whose graph lies in Γ and

$$g(x,\mu(x)) + \alpha\int J^*(x')t(dx'|x,\mu(x)) = \inf_{u\in\Gamma_x}\{g(x,u) + \alpha\int J^*(x')t(dx'|x,u)\}.$$

Then $\pi=(\mu,\mu,\ldots)$ is optimal by Theorems 5.1 and 5.5.

If $\pi=(\mu,\mu,\ldots)$ is an optimal nonrandomized stationary policy, then

$$T_\mu J^* = T_\mu J_\mu = J_\mu = J^* = TJ^*,$$

so $\mu(x)$ achieves the above infimum for every x. QED

Under (N) we can use Theorem 5.3 to obtain a necessary and sufficient condition for a stationary policy to be optimal. This condition is not as useful as that of Theorem 5.5, however, since it cannot be used to construct a stationary optimal policy in the manner of Corollary 5.5.1.

<u>Theorem 5.6</u> (N)(D)

Let $\bar{\pi}=(\bar{\mu},\bar{\mu},\ldots)$ and $\pi=(\mu,\mu,\ldots)$ be stationary policies in (DM) and (SM) respectively. The policy $\bar{\pi}$ is optimal if and only if $\bar{J}_{\bar{\mu}}=\bar{T}\bar{J}_{\bar{\mu}}$. The policy π is optimal if and only if $J_\mu=TJ_\mu$.

Proof:

Again the proof for (DM) can be found in [4] or [3] but is given here.

If $\bar{\pi}$ is optimal, $\bar{J}_{\bar{\mu}}=\bar{J}^*$. By Theorem 5.1,

$$\bar{J}_{\bar{\mu}} = \bar{J}^* = \bar{T}\bar{J}^* = \bar{T}\bar{J}_{\bar{\mu}}.$$

Conversely, if $\bar{J}_{\bar{\mu}}=\bar{T}\bar{J}_{\bar{\mu}}$, then by Theorem 5.3, $\bar{J}_{\bar{\mu}}\le\bar{J}^*$ and $\bar{\pi}$ is optimal.

If π is optimal, $J_\mu = TJ_\mu$ by the (SM) part of Theorem 5.1. The converse is more difficult, since the (SM) part of Theorem 5.3 cannot be invoked without knowing that J_μ is lower semianalytic. Let $\bar{\pi}=(\bar{\mu},\bar{\mu},\ldots)$ correspond to $\pi=(\mu,\mu,\ldots)$, so that $\bar{J}_{\bar{\mu}}(p)=\int J_\mu dp$ for every $p\varepsilon P(S)$. Then for fixed $p\varepsilon P(S)$ and $q\varepsilon P(\Gamma)_p$,

$$\int g\,dq + \alpha\bar{J}_{\bar{\mu}}[\bar{f}(q)] = \int_{SC}[g(x,u) + \alpha\int_S J_\mu(x')t(dx'|x,u)]q(d(x,u))$$

$$\geq \int_S \inf_{u \varepsilon \Gamma_x} [g(x,u) + \alpha \int_S J_\mu(x')t(dx'|x,u)]p(dx)$$

provided the integrand is universally measurable

$$= \int_S (TJ_\mu)dp$$

$$= \int_S J_\mu dp, \text{ so the integrand is universally measurable,}$$

$$= \bar{J}_{\bar{\mu}}(p).$$

Infimizing the left hand side over $q \varepsilon P(\Gamma)_p$, we see that

$$\bar{T}\bar{J}_{\bar{\mu}} \geq \bar{J}_{\bar{\mu}} = \bar{T}_{\bar{\mu}}\bar{J}_{\bar{\mu}}.$$

The reverse inequality always holds, and by the proof already given for (DM), $\bar{\pi}$ is optimal. Then π is optimal by Corollary 4.o.1. QED

Section 5. Convergence of the dynamic programming algorithm

Definition 5.1 The dynamic programming algorithm is defined recursively for (DM) and (SM) by

$$\bar{J}_o = 0,$$
$$\bar{J}_{k+1} = \bar{T}\bar{J}_k, \ k=0,1,\ldots,$$
$$J_o = 0,$$
$$J_{k+1} = TJ_k, \ k=0,1,\ldots.$$

We saw in Chapter 3 that this algorithm generates the k-stage optimal cost function J_k^*. For simplicity of notation, we suppress the * here. At present we are concerned with the infinite horizon cases and the possibility that J_k may converge to J^* as $k \to \infty$.

Under (P), $\bar{J}_0 \leq \bar{J}_1$ and so $\bar{J}_1 = \bar{T}\bar{J}_0 \leq \bar{T}\bar{J}_1 = \bar{J}_2$. Continuing, we see that $\bar{J}_k$ is an increasing sequence of functions, and so $\bar{J}_\infty = \lim_k \bar{J}_k$ exists and takes values in $[0,+\infty]$. Under (N), $\bar{J}_k$ is a decreasing sequence of functions and $\bar{J}_\infty$ exists, taking values in $[-\infty,0]$. Under (D), if $-c \leq \bar{J} \leq c < \infty$,

$$\begin{aligned} 0 &\leq \bar{J} + c \\ &\leq b + \bar{T}(\bar{J} + c) \\ &= b + \alpha c + \bar{T}\bar{J} \\ &\leq b + \bar{T}(b + \alpha c + \bar{T}\bar{J}) \\ &\leq b + \alpha b + \alpha^2 c + \bar{T}^2\bar{J}, \end{aligned}$$

and in general,

$$0 \leq \lim_{k\to\infty} [b\sum_{j=0}^{k-1}\alpha^j + \alpha^k c + \bar{T}^k\bar{J}] = b/(1-\alpha) + \lim_{k\to\infty} \bar{T}^k\bar{J}.$$

Similarly,

$$0 \geq -b/(1-\alpha) + \lim_{k\to\infty} \bar{T}^k\bar{J}.$$

Therefore, $\lim_k \bar{T}^k\bar{J}$ exists and takes values in $[-b/(1-\alpha), b/(1-\alpha)]$. In particular, $\bar{J}_\infty$ exists and takes values in $[-b/(1-\alpha), b/(1-\alpha)]$.

The same arguments can be used to establish the existence of $J_\infty = \lim_k J_k$. Under (P), $J_\infty : S \to [0,+\infty]$; under (N), $J_\infty : S \to [-\infty,0]$; and under (D), $\lim_k T^k J : S \to [-b/(1-\alpha), b/(1-\alpha)]$, where $J : S \to [-c,c]$, $c<\infty$, is lower semianalytic. Note that in every case, the lower level sets

$$\{J_\infty \leq c\} = \bigcap_{n=1}^{\infty} \bigcup_{K=1}^{\infty} \bigcap_{k=K}^{\infty} \{J_k \leq c + 1/n\}$$

are analytic by Theorem A.2, so J_∞ is lower semianalytic.

Lemma 5.1 (P)(N)(D)

For every $p \in P(S)$, $\bar{J}_k(p)=\int J_k(x)p(dx)$, $k=0,1,\ldots$ and $k=\infty$.

Proof:

Applying Theorem 4.4, use induction to prove the lemma for $k=0,1,\ldots$. When $k=\infty$, the lemma follows from the monotone convergence theorem under (P) or (N) and the bounded convergence theorem under (D). QED

Theorem 5.7 (N)(D)

We have

$$\bar{J}_\infty = \bar{J}^*, \quad J_\infty = J^*.$$

Indeed, under (D) the dynamic programming algorithm can be initiated from any $\bar{J}:P(S)\to[-c,c]$, $c<\infty$, or lower semianalytic $J:S\to[-c,c]$, $c<\infty$, and converges uniformly, i.e.

$$\lim_{k\to\infty} \sup_{p\in P(S)}|(\bar{T}^k\bar{J})(p) - \bar{J}^*(p)| = 0,$$
$$\lim_{k\to\infty} \sup_{x\in S}|(T^kJ)(x) - J^*(x)| = 0.$$

Proof:

The theorem for (DM) follows from [4], Proposition 1 and Lemma 1 or [3], Chapter 6, Proposition 3, and Chapter 7, Proposition 4. By Lemma 5.1, $J_k(x)=\bar{J}_k(p_x)$, $k=0,1,\ldots$ and $k=\infty$, so the theorem holds for (SM) under (N) as well. Under (D), define $\bar{J}(p)=\int J dp$ and use Theorem 4.4 and the (DM) result. QED

The case (D) is the best suited for computational procedures. The machinery developed thus far can be applied to the proof of [3], Chapter 6, Proposition 4, to show the validity for (SM) of the error bounds given there. We state the theorem for (SM). The analogous result is true for (DM).

Theorem 5.8 (D)

Let $J:S \to [-c,c]$, $c<+\infty$, be lower semianalytic. Then for all $x \in S$ and $k=0,1,\ldots,$

$$(T^kJ)(x)+c_k \le (T^{k+1}J)(x)+c_{k+1} \le J^*(x) \le (T^{k+1}J)(x)+\bar{c}_{k+1} \le (T^kJ)(x)+\bar{c}_k,$$

where,

$$c_k = [\alpha/(1-\alpha)]\inf_{x\in S}[(T^kJ)(x) - (T^{k-1}J)(x)],$$
$$\bar{c}_k = [\alpha/(1-\alpha)]\sup_{x\in S}[(T^kJ)(x) - (T^{k-1}J)(x)].$$

Without further assumptions we have only the following weaker results concerning convergence of the dynamic programming algorithm under (P).

Theorem 5.9 (P)

It holds that

$$\bar{J}_\infty \le \bar{J}^*, \quad J_\infty \le J^*.$$

Furthermore, the following statements are equivalent:

(a) $\bar{J}_\infty = \bar{T}\bar{J}_\infty$,

(b) $\bar{J}_\infty = \bar{J}^*$,

(c) $J_\infty = TJ_\infty$,

(d) $J_\infty = J^*$.

Proof:

By [4], Proposition 10, it holds that $\bar{J}_\infty \le \bar{J}^*$. By the same reference or by [3], Chapter 7, Proposition 4, (a) and (b) are equivalent. By Lemma 5.1 and Theorem 4.6, it holds that $J_\infty(x) \le J^*(x)$. Conditions (a) and (c) are equivalent by Lemma 5.1 and Theorem 4.4. Conditions (b) and (d) are equivalent by Lemma 5.1 and Theorem 4.6. QED

<u>Theorem 5.10</u> (P)(D)

Assume that the sets

$$U_k(x,c) = \{u \in \Gamma_x : g(x,u) + \alpha \int J_k(x')t(dx'|x,u) \leq c\}$$

are compact subsets of C for every $x \in S$, $c \in R$, and for all k greater than some integer $\bar{k}$. Then the equivalent conditions of Theorem 5.9 hold and there exists a nonrandomized stationary optimal policy.

Proof:

Use the proof of Proposition 13 of Chapter 6 of [3] and Corollary 5.5.1.

QED

The next two corollaries give conditions under which the assumptions of the theorem are satisfied. Note that under (D) the only new result in the theorem and corollaries is the existence of a stationary optimal policy. The equivalent conditions of Theorem 5.9 always hold under (D).

<u>Corollary 5.10.1</u> (P)(D)

Assume the set Γ_x is finite for each $x \in S$. Then the equivalent conditions of Theorem 5.9 hold and there exists a nonrandomized stationary optimal policy.

<u>Corollary 5.10.2</u> (P)(D)

Let (3.16) - (3.19) hold. Then the equivalent conditions of Theorem 5.9 hold and there exists a Borel measurable nonrandomized stationary optimal policy.

Proof:

Theorem 3.4 establishes that the functions J_n are lower semicontinuous and bounded below on S. By the proof of Theorem 3.4, the functions

$$H_n(x,u) = g(x,u) + \alpha \int J_n(x')t(dx'|x,u) \text{ if } (x,u) \in \Gamma;$$

AD-A038 337 ILLINOIS UNIV AT URBANA-CHAMPAIGN COORDINATED SCIENCE LAB F/G 12/2
DYNAMIC PROGRAMMING IN COMPLETE SEPARABLE SPACES.(U)
JAN 77 S E SHREVE DAAB07-72-C-0259
UNCLASSIFIED R-755 NL

2 OF 2
AD
A038337

END
DATE
FILMED
5 - 77

$= +\infty$ otherwise;

are lower semicontinuous. For $c\in R$ and n fixed, condition (3.17) implies the existence of j for which

$$\{(x,u)\in SC:\ H_n(x,u)\le c\}\subset \Gamma^j.$$

This lower level set is closed by the lower semicontinuity of H_n and this implies the compactness of $U_n(x,c)$ for each $x\in S$. Theorem 5.10 can now be invoked and it remains only to prove that the nonrandomized stationary optimal policy whose existence is guaranteed by that theorem can be chosen to be Borel measurable. This will follow from Theorems 3.4 and 5.5 if $J_\infty = J^*$ can be shown to be lower semicontinuous. Under (P)

$$J_\infty = \sup_n J_n,$$

while under (D)

$$J_\infty = \sup_n\{J_n - b\sum_{k=n+1}^{\infty}\alpha^k\}.$$

In either case, J_∞ is lower semicontinuous by Lemma 2.5(b). QED

Theorem 5.11 (P)(D)

Suppose either that the set Γ_x is finite for each $x\in S$, or else (3.16) - (3.19) hold. Let $\mu_k: S\to C$ be a sequence of universally measurable functions such that $\mu_k(x)\in\Gamma_x$ for every x, and $T_{\mu_k}J_k = TJ_k$, $k=0,1,\ldots$. Assume that $\{x: J^*(x)<\infty\}$ can be partitioned into countably many disjoint universally measurable sets $B_1, B_2, \ldots$ such that on each B_j a subsequence of $\{\mu_k\}$ converges to a function μ'. Then μ' can be extended to a universally measurable function μ on S such that $\mu(x)\in\Gamma_x$ for every x, and $\pi=(\mu,\mu,\ldots)$ is optimal.

Proof:

Clearly μ' is universally measurable. Suppose first that (3.16) - (3.19) hold. From the proof of Theorem 3.4 we see that the functions

$$H_n(x,u) = g(x,u) + \alpha\int J_n(x')t(dx'|x,u) \text{ if } (x,u)\varepsilon\Gamma;$$
$$= +\infty \text{ otherwise;}$$

are lower semicontinuous, n=0,1,.... Choose $x\varepsilon B_j$ and let N_j be an infinite subset of the positive integers such that $\{\mu_k\}_{k\varepsilon N_j}$ converges to μ' on B_j. Then for n fixed,

$$\begin{aligned}
(5.1) \qquad \infty > J^*(x) &= \lim_{k\varepsilon N_j} J_{k+1}(x) \\
&= \lim_{k\varepsilon N_j}[g(x,\mu_k(x)) + \alpha\int J_k(x')t(dx'|x,\mu_k(x))] \\
&\geq \lim\sup_{k\varepsilon N_j}[g(x,\mu_k(x)) + \alpha\int J_n(x')t(dx'|x,\mu_k(x))] \\
&= \lim\sup_{k\varepsilon N_j} H_n(x,\mu_k(x)) \\
&\geq H_n(x,\mu'(x)) \text{ by the lower semicontinuity of } H_n \\
&= (T_{\mu'}J_n)(x) \\
&\geq (TJ_n)(x) \\
&= J_{n+1}(x).
\end{aligned}$$

This implies $(x,\mu'(x))\varepsilon\Gamma$. Use von Neumann's Lemma (Theorem A.6) to extend μ' to a universally measurable function μ from S to C satisfying $(x,\mu(x))\varepsilon\Gamma$ for every x. Now let $n\to\infty$ in (5.1) and from Lemma 3.4(b) or (c) and Theorem 5.10, conclude $TJ^*=T_\mu J^*$. Apply Theorem 5.5.

If instead of (3.16) - (3.19), we have that Γ_x is finite for each $x\varepsilon S$, the argument is still valid if we use the finiteness of Γ_x to show

$$\lim \sup_{k\varepsilon N_j} H_n(x,\mu_k(x)) = H_n(x,\mu'(x))$$

in establishing (5.1). QED

Section 4. Existence of optimal and ϵ-optimal policies

Theorem 5.12 (P)(D)

For each $\epsilon>0$, there exists an ϵ-optimal nonrandomized Markov policy for (SM), and if $\alpha<1$, it can be taken to be stationary. If for each $x\varepsilon S$ there exists a policy optimal at x for (SM), then an optimal nonrandomized stationary policy exists.

Proof:

Choose $\epsilon>0$ and $\epsilon_k>0$ such that $\sum_{k=0}^{\infty}\alpha^k\epsilon_k=\epsilon$. If $\alpha<1$, let $\epsilon_k=(1-\alpha)$ for every k. By Theorem 2.5, there are universally measurable functions $\mu_k:S\rightarrow C$, k=0,1,..., such that $\mu_k(x)\varepsilon\Gamma_x$ for every x and

$$T_{\mu_k}J^* \leq J^* + \epsilon_k.$$

If $\alpha<1$, we can choose all the μ_k's identical. Then

$$(T_{\mu_{k-1}}T_{\mu_k})(J^*) \leq (T_{\mu_{k-1}}J^*) + \alpha\epsilon_k \leq J^* + \epsilon_{k-1} + \alpha\epsilon_k.$$

Continuing this process, we have

$$(T_{\mu_0}T_{\mu_1}\cdots T_{\mu_k})(J_0) \leq (T_{\mu_0}T_{\mu_1}\cdots T_{\mu_k})(J^*) \leq J^* + \sum_{j=0}^{k}\alpha^j\epsilon_j \leq J^* + \epsilon,$$

and letting $k\rightarrow\infty$, we obtain

$$J_\pi \leq J^* + \epsilon,$$

where $\pi=(\mu_0,\mu_1,\ldots)$. This proves the first part of the theorem.

Suppose $\pi'=(\mu_0',\mu_1',\ldots)\in\Pi'$ is a policy for (SM) which is optimal at $x\in S$. By Theorem 3.1, there is a $\pi=(\mu_0,\mu_1,\ldots)\in\Pi$ such that $J_\pi(x) = J_{\pi'}(x) = J^*(x)$ and

$$\int_{\underline{S}} \mu_0(\underline{C}|x_0)p_x(dx_0) = \int_{\underline{S}} \mu_0'(\underline{C}|x_0)p_x(dx_0)$$

for every $\underline{S}\in\mathcal{B}_S$, $\underline{C}\in\mathcal{B}_C$.

Therefore

$$(T_{\mu_0}J^*)(x) = (T_{\mu_0'}J^*)(x).$$

Now

$$\begin{aligned}
J^*(x) &= \lim_k (T_{\mu_0}T_{\mu_1}\ldots T_{\mu_k})(J^*)(x) \\
&= T_{\mu_0}[\lim_k (T_{\mu_1}\ldots T_{\mu_k})(J^*)](x) \quad \text{by Lemma 3.4(b) or (c)} \\
&\geq (T_{\mu_0}J^*)(x) \\
&\geq (TJ^*)(x) \\
&= J^*(x).
\end{aligned}$$

Consequently, we have $(T_{\mu_0'}J^*)(x)=(TJ^*)(x)$. This implies that for each x, the infimum in the expression

$$\inf_{u\in\Gamma_x}[g(x,u) + \alpha\int J^*(x')t(dx'|x,u)]$$

is achieved. The conclusion follows from Corollary 5.5.1. QED

<u>Theorem 5.13</u> (N)

For each $\epsilon>0$, there exists an ϵ-optimal nonrandomized semi-Markov policy. If for each $x\in S$, there exists a policy optimal at x for (SM), then there exists a

semi-Markov (randomized) optimal policy.

Proof:

Under (N), $J_k \downarrow J^*$ (Theorem 5.7), and so given $\epsilon>0$, the analytically measurable sets

$$A_k = \{x: J_k(x)<J^*(x)+(\epsilon/2) \text{ if } J^*(x)>-\infty,\ J_k(x)\leq-(1/\epsilon)-(\epsilon/2) \text{ if } J^*(x)=-\infty\}$$

converge up to S. Use Corollary 3.2.2 to define for each positive integer k a nonrandomized semi-Markov policy $\pi^k=(\mu_0^k,\ldots,\mu_{k-1}^k,\mu,\mu,\ldots)$ such that

$$\begin{aligned} J_{\pi^k}(x) &\leq J_k(x) + \epsilon/2 \text{ if } J_k(x)>-\infty; \\ &\leq -1/\epsilon \text{ otherwise;} \end{aligned}$$

for every $x\epsilon S$. Then for $x\epsilon A_k$,

$$\begin{aligned} J_{\pi^k}(x) &\leq J^*(x) + \epsilon \text{ if } J^*(x)>-\infty; \\ &\leq -1/\epsilon \text{ otherwise.} \end{aligned}$$

The policy π defined to be π^k when the initial state is in A_k, but not in A_j for any $j<k$, is ϵ-optimal, semi-Markov and nonrandomized.

Suppose now that for each $x\epsilon S$, there exists a policy optimal at x for (SM). Let $G:P(S)P(SC)\ldots\rightarrow[-\infty,0]$ be as defined in Theorem 4.5. Then for each p_x, $x\epsilon S$, there is an admissible sequence $(p_x,q_0^x,q_1^x,\ldots)\,\epsilon\,\Delta$ which achieves the infimum in

$$\inf_{(p_0,q_0,\ldots)\epsilon\Delta,\,p_0=p_x} G(p_0,q_0,\ldots)$$

(Theorem 4.1 and Corollary 4.2.1).

The set $\{p_x: x \in S\}P(SC)P(SC)\ldots$ is analytic, so G is lower semianalytic on this set. There is a universally measurable selector $\varphi: S \to P(SC)P(SC)\ldots$ such that $(p_x, \varphi(x)) \in \Delta$ and

$$G(p_x, \varphi(x)) = J^*(x)$$

for every $x \in S$ (Theorems 2.5 and B.7). Denote

$$\varphi(x) = (q_o(\cdot|x), q_1(\cdot|x), \ldots).$$

By Theorem C.2, for each x and k, $q_k(\cdot|x)$ has a decomposition into its marginal $p_k(\cdot|x)$ on S_k and a universally measurable stochastic kernel $\mu_k(du_k|x,x_k)$ on C_k given S_oS_k. The policy $\pi = (\mu_o, \mu_1, \ldots)$ is semi-Markov and optimal. QED

Although randomized policies are intuitively inferior and avoided in practice, under (N) as posed here, they cannot be disregarded, as the following example demonstrates.

Example 5.1 (Petersburg Paradox)

Let $S=\{0,1,2,\ldots\}$, $C=\{0,1\}$, $W=\{w\}$,

$$g(x,u) = -2^x \text{ if } x \neq 0,\ u=0; \qquad = 0 \text{ otherwise};$$

$$f(x,u,w) = x+1 \text{ if } u=1,\ x \neq 0; \qquad = 0 \text{ otherwise}.$$

Beginning in state 1, any nonrandomized policy increases the state by one for finitely many, say k, moves at no cost and then jumps to zero at a cost of -2^{k+1}, where the state remains at no further cost. Thus $J^*(1)=-\infty$, but this value is not achieved by any nonrandomized policy. On the other hand, the randomized policy which advances with probability 1/2 when the state x is nonzero yields an expected cost of $-\infty$.

The one-stage cost g in Example 5.1 is unbounded, but by a slight modification, an example can be constructed in which g is bounded and the only optimal policies are randomized. If one stipulates that J^* must be finite, it may be possible to restrict attention to nonrandomized policies in Theorem 5.13. This is an unsolved problem.

CHAPTER 6

GENERALIZATIONS OF RESULTS

Section 1. The nonstationary model

The model of Definition 3.1 is said to be stationary, i.e. the state, control and disturbance spaces, constraint set, cost and system function, and disturbance kernel are independent of the stage. A nonstationary model can be reduced to a stationary one by the technique described in Chapter 6, Section 7, of Bertsekas [3] or Section 8 of Schael [32]. If the state spaces $S_o,\ldots,S_{N-1}$ and the control spaces $C_o,\ldots,C_{N-1}$ are nonempty Borel spaces, the constraint sets Γ_k are analytic subsets of $S_k C_k$ satisfying $(\Gamma_k)_{x_k} \neq \emptyset$ for every $x_k \in S_k$, $k=0,\ldots,N-1$, the cost functions $g_k: S_k C_k \to R^*$ are lower semianalytic, $k=0,\ldots,N-1$, and the state transition kernels $t_k(dx_{k+1}|x_k,u_k)$ are Borel measurable, $k=0,\ldots,N-2$, then we define new state and control spaces

$$S = \{(x,k): x \in S_k\},$$

$$C = \{(u,k): u \in C_k\},$$

with the metric on S

$$\begin{aligned} d[(x,i),(y,j)] &= d_i(x,y) \text{ if } i=j; \\ &= \text{diameter}(S_i) + \text{diameter}(S_j) \text{ if } i \neq j; \end{aligned}$$

where d_i is a metric on S_i consistent with its topology. If $N<\infty$, we also include in S an isolated point T. A similar metric is defined on C.

Define the constraint set $\Gamma=\{((x,k),(u,k)):\ 0 \leq k \leq N-1,\ (x,u) \in \Gamma_k\}$, the cost function

$$g((x,i),(u,j)) = g_i(x,u) \text{ if } i=j;$$
$$= 0 \text{ if } i\neq j;$$

and the state transition probability

$$t(\underline{S}|(x,i),(u,j)) = t_i(\{x': (x',i+1)\in\underline{S}\}|x,u) \text{ if } i=j\neq N-1;$$
$$= p(\underline{S}) \text{ otherwise};$$

where $p=p_T$ if $N<\infty$ and is any fixed probability measure on S if $N=\infty$.

Defined in this manner, S,C,Γ,g and t satisfy the assumptions of the stationary stochastic model (SM). Conditions on the g_k's such as nonnegativity or uniform boundedness in k result in the same conditions on g. Identifying S_k with $\{(x,k): x\in S_k\}$ and C_k with $\{(u,k): u\in C_k\}$, there is a clear correspondence between the stationary and nonstationary models. The system in the stationary model moves from the set $\{(x,k): x\in S_k\}$ to the set $\{(x,k+1): x\in S_{k+1}\}$ with probability one at each stage. If $N<\infty$, the system moves from $\{(x,N-1): x\in S_{N-1}\}$ to T and remains there at no further cost. Universally measurable policies in the two models correspond and result in the same expected cost.

Because such a reduction is possible, results already proved for the stationary model with either a finite or an infinite horizon have immediate counterparts for the nonstationary model. An illustration of this is the nonstationary optimality equation (Theorem 5.1).

Theorem 6.1 (P)(N)(D)

Let $S_o,S_1,\ldots;$ $C_o,C_1,\ldots;$ $\Gamma_o,\Gamma_1,\ldots;$ $g_o,g_1,\ldots;$ and $t_o,t_1,\ldots$ be as described above and suppose either

(P) $0 \leq g_k(x,u)$ for every $(x,u)\in S_kC_k$, $k=0,1,\ldots,$

(N) $0 \geq g_k(x,u)$ for every $(x,u) \in S_k C_k$, $k=0,1,\ldots$, or

(D) $-b \leq g_k(x,u) \leq b < \infty$ for every $(x,u) \in S_k C_k$, $k=0,1,\ldots$, and $0 < \alpha < 1$.

Let $J^*(x,k)$ be the optimal cost associated with state x in space S_k. Then for each k, $J^*(x,k)$ is lower semianalytic on S_k and satisfies

$$J^*(x,k) = \inf_{u \in (\Gamma_k)_x} [g(x,u) + \alpha \int_{S_{k+1}} J^*(x',k+1) t_k(dx' | x,u)],\ k=0,1,\ldots.$$

We will henceforth apply nonstationary results and reference only their stationary counterparts. We get an easy theorem for the stationary (SM) in the following manner.

Definition 6.1 Given $p \in P(S)$ and $\epsilon > 0$, a policy π in (SM) is weakly p-ϵ-optimal provided

$$\int J_{N,\pi}(x)p(dx) \leq \int J_N^*(x)p(dx) + \epsilon \text{ if } \int J_N^*(x)p(dx) > -\infty;$$
$$\leq -1/\epsilon \text{ if } \int J_N^*(x)p(dx) = -\infty.$$

A policy π in (SM) is p-optimal provided $p\{x: J_{N,\pi}(x) = J_N^*(x)\} = 1$.

Theorem 6.2 $(F^+)(F^-)(P)(N)(D)$

Given $\epsilon > 0$, there is a set of nonrandomized policies

$$\pi(p) = (\mu_0(du_0 | p;x_0),\ldots,\mu_{N-1}(du_{N-1} | p;x_{N-1}))$$

such that for $k=0,1,\ldots,N-1$, μ_k is universally measurable in $(p;x_k)$ and $\pi(p)$ is weakly p-ϵ-optimal for every $p \in P(S)$.

Proof:

For (F^+), (P) and (D), stronger results have already been proved (Corollary 3.2.2 and Theorem 5.12). Consider now the nonstationary problem where $S_0 = P(S)$, $C_0 = \{u_0\}$, $\Gamma_0 = S_0 C_0$, $g_0(p,u_0) = 0$ for every (p,u_0), and $t_0(\cdot | p,u_0) = p$. The subsequent state spaces, control spaces, constraint sets,

cost functions, and transition probabilities are S,C,Γ,g, and t of the stationary model respectively. If J_N^* is the optimal cost function in the stationary model, then for $p\epsilon S_o$, the optimal cost at p in the nonstationary model is $\int J_N^*(x)p(dx)$ (Corollary 3.2.4 or Corollary 4.6.2).

There exists an ϵ-optimal nonrandomized semi-Markov policy in the nonstationary model (Corollary 3.2.2 or Theorem 5.13) and, omitting the first function in this policy (which is identically u_o), we obtain a set of weakly p-ϵ-optimal policies in the stationary model. QED

Section 2. The imperfect state information model

Definition 6.1 An imperfect state information stochastic decision model (ISI) is the ten-tuple $(S,C,(\Gamma_o,\ldots,\Gamma_{N-1}),Z,\alpha,g,t,s_o,s,N)$ described below.

S,C,α,g,t: State space, control space, discount factor, one-stage cost function, and state transition kernel as defined in Definition 3.1, (3.2) and (3.1).

Z: Observation space. A nonempty Borel space.

Γ_k: k-th constraint set. An analytic subset of I_kC, where $I_k=ZC\ldots CZ$, the Z appearing $k+1$ times and the C appearing k times. An element of I_k is called a k-th information vector. The constraint sets Γ_k satisfy $(\Gamma_k)_{i_k}\neq\emptyset$ for every $i_k\epsilon I_k$.

s_o: Initial observation kernel. A Borel measurable stochastic kernel on Z given S.

s: Observation kernel. A Borel measurable stochastic kernel on Z given CS.

N: Horizon. A positive integer or $+\infty$.

The system moves stochastically from state x_k to state x_{k+1} via the state transition kernel $t(dx_{k+1}|x_k,u_k)$ and generates cost at each stage of $g(x_k,u_k)$. The observation z_{k+1} is stochastically generated via the observation kernel $s(dz_{k+1}|u_k,x_{k+1})$ and added to the past observations and controls $(z_0,u_0,\ldots,z_k,u_k)$ to form the $(k+1)$-st information vector $i_{k+1}=(z_0,u_0,\ldots,z_k,u_k,z_{k+1})$. The first information vector $i_0=(z_0)$ is generated by the initial observation kernel $s_0(dz_0|x_0)$, and the initial state x_0 has some given initial distribution p. The goal is to choose u_k dependent on the k-th information vector i_k so as to minimize

$$E\{\sum_{k=0}^{N-1}\alpha^k g(x_k,u_k)\}.$$

In what follows, our notation will generally indicate a finite N. If N is infinite, the appropriate interpretation is required.

Definition 6.2 A policy in (ISI) is a sequence $\pi=(\mu_0,\ldots,\mu_{N-1})$ such that for each k, $\mu_k(du_k|p;i_k)$ is a universally measurable stochastic kernel on C given $P(S)I_k$ satisfying

$$\mu_k((\Gamma_k)_{i_k}|p;i_k) = 1$$

for every $(p;i_k)$. If for each p, k and i_k, $\mu_k(\cdot|p;i_k)$ assigns mass one to some point in C, π is nonrandomized. A policy is said to be Borel measurable if all its component stochastic kernels are.

The concepts of Markov and semi-Markov policies are of no use in (ISI), since the initial distribution, past observations and past controls are of genuine value in estimating the current state. Thus we expect policies to depend on the initial distribution p and the total information vector. In this chapter, Π will denote the set of all policies in (ISI).

Just as we denote the set of all sequences of the form $(z_o,u_o,\ldots,u_{k-1},z_k)\varepsilon ZC\ldots CZ$ by I_k and call these sequences the k-th information vectors, we find it notationally convenient to denote the set of all sequences of the form $(x_o,z_o,u_o,\ldots,x_k,z_k,u_k)\varepsilon SZC\ldots SZC$ by H_k and call these sequences the k-th <u>history vectors</u>. Except for u_k, the k-th information vector is that portion of the k-th history vector known to the controller.

Given $p\varepsilon P(S)$ and $\pi=(\mu_o,\ldots,\mu_{N-1})\varepsilon\Pi$, there is a sequence of consistent probability measures $P_k(\pi,p)$ generated on H_k, $k=0,\ldots,N-1$, defined on measurable rectangles by

(6.1) $$P_k(\pi,p)\{\underline{S}_o\underline{Z}_o\underline{C}_o\ldots\underline{S}_k\underline{Z}_k\underline{C}_k\}=\int_{\underline{S}_o}\int_{\underline{Z}_o}\int_{\underline{C}_o}\cdots\int_{\underline{S}_k}\int_{\underline{Z}_k}\mu_k(\underline{C}_k|p;z_o,u_o,\ldots,u_{k-1},z_k)$$

$$s(dz_k|u_{k-1},x_k)t(dx_k|u_{k-1},x_{k-1})\cdots\mu_o(du_o|p;z_o)s_o(dz_o|x_o)p(dx_o).$$

<u>Definition 6.3</u>

Given $p\varepsilon P(S)$, a policy $\pi=(\mu_o,\ldots,\mu_{N-1})\varepsilon\Pi$, and a positive integer $K\leq N$, the <u>K-stage cost function corresponding to π</u> at p is

(6.2) $$J_{K,\pi}(p) = \sum_{k=0}^{K-1}\alpha^k\int g(x_k,u_k)dP_k(\pi,p).$$

The <u>cost function corresponding to π</u> is $J_{N,\pi}$. If $N=\infty$, we impose either condition (P), (N) or (D) of Chapter 4, Section 1, on the model to ensure that the sum in (6.2) is a well-defined extended real number. If $N<\infty$, we will assume either (F^-):

$$\int g^+(x_k,u_k)dP_k(\pi,p) < \infty,\ k=0,\ldots,N-1,$$

for every $\pi\varepsilon\Pi$ and $p\varepsilon P(S)$ or (F^+):

$$\int g^-(x_k,u_k)dP_k(\pi,p) < \infty,\ k=0,\ldots,N-1,$$

for every $\pi\in\Pi$ and $p\in P(S)$.

The *optimal cost function* at p is

$$J_N^*(p) = \inf_{\pi\in\Pi} J_{N,\pi}(p). \tag{6.3}$$

The concepts of *optimality at p*, *optimality*, *ϵ-optimality at p* and *ϵ-optimality* of policies are the same as those given in Definition 3.6.

To aid in the analysis of (ISI) we introduce the idea of a statistic sufficient for control, which allows us to revert to a nonstationary perfect state information model.

Definition 6.4 A *statistic* is a sequence $\Sigma=(\sigma_0,\ldots,\sigma_{N-1})$ of Borel measurable functions $\sigma_k: P(S)I_k \to Y_k$, where Y_k is a Borel space, $k=0,\ldots,N-1$. A statistic $\Sigma=(\sigma_0,\ldots,\sigma_{N-1})$ is *sufficient for control* provided:

(a) If for some k, $p,p'\in P(S)$, and $i_k,i_k'\in I_k$, we have $\sigma_k(p;i_k)=\sigma_k(p';i_k')$, then $(\Gamma_k)_{i_k}=(\Gamma_k)_{i_k'}$;

(b) There exist Borel measurable stochastic kernels $\hat{t}(dy_{k+1}|y_k,u_k)$ on Y_{k+1} given Y_kC such that for every $p\in P(S)$, $\pi\in\Pi$, $\underline{Y}_{k+1}\in\mathcal{B}_{Y_{k+1}}$,

$$P_{k+1}(\pi,p)\{\sigma_{k+1}(p;\cdot)\in\underline{Y}_{k+1} \mid \sigma_k(p;i_k)=y_k,\ u_k=u\} = \hat{t}_k(\underline{Y}_{k+1}|y_k,u)^{1} \tag{6.4}$$

$P_{k+1}(\pi,p)$-almost surely,[2] $k=0,\ldots,N-2$;

---

[1] We use the notation $\sigma_{k+1}(p;\cdot)\in\underline{Y}_{k+1}$ to indicate the set $\{(x_0,z_0,u_0,\ldots,x_{k+1},z_{k+1},u_{k+1})\in H_{k+1} \mid \sigma_{k+1}(p;(z_0,u_0,\ldots,u_k,z_{k+1}))\in\underline{Y}_{k+1}\}$. Whenever a subset depending on some of the components of a Cartesian product is considered, this type of notation will be employed.

[2] In this context "$P_{k+1}(\pi,p)$-almost surely" means that the set $\{(x_0,z_0,u_0,\ldots,x_{k+1},z_{k+1},u_{k+1})\in H_{k+1} \mid (6.4) \text{ holds when } y_k=\sigma_k(p;i_k), u=u_k\}$ has $P_{k+1}(\pi,p)$ measure one.

(c) There are lower semianalytic functions $\hat{g}_k: Y_k C \rightarrow R^*$ satisfying for every $p \in P(S)$ and $\pi \in \Pi$,

$$E\{g(x_k,u_k) | \sigma_k(p;\cdot)=y_k, u_k=u\} = \hat{g}_k(y_k,u) \tag{6.5}$$

$P_k(\pi,p)$-almost surely, $k=0,\ldots,N-1$. The expectation is with respect to $P_k(\pi,p)$.

It is no additional restriction to assume that under (P), $\hat{g}_k \geq 0$, $k=0,\ldots,N-1$; under (N), $\hat{g}_k \leq 0$, $k=0,\ldots,N-1$; and under (D), $-b \leq \hat{g}_k \leq b$, $k=0,\ldots,N-1$. We make this assumption.

Condition (a) of Definition 6.4 guarantees that the constraint set section $(\Gamma_k)_{i_k}$ can be recovered from $\sigma_k(p;i_k)$. Define for $k=0,\ldots,N-1$,

$$\hat{\Gamma}_k = \{(y_k,u): \text{ if } y_k=\sigma_k(p;i_k), \text{ then } (i_k,u) \in \Gamma_k\}. \tag{6.6}$$

Then by (a), for any p, $(\hat{\Gamma}_k)_{y_k} \neq \emptyset$ for every y_k and

$$\Gamma_k = \{(i_k,u): (\sigma_k(p;i_k),u) \in \hat{\Gamma}_k\}. \tag{6.7}$$

If $\Gamma_k = I_k C$, condition (a) is satisfied and $\hat{\Gamma}_k = Y_k C$. This is the case of no control constraint.

Condition (b) guarantees that the distribution of σ_{k+1} depends only on the values of σ_k and u_k. This is necessary in order that the y_k's form the states of a stochastic decision model.

Condition (c) guarantees that the cost corresponding to a policy can be computed from the distributions induced on the (y_k,u_k) pairs.

<u>Definition 6.5</u> Let $(S,C,(\Gamma_0,\ldots,\Gamma_{N-1}),Z,\alpha,g,t,s_0,s,N)$ be an imperfect state information stochastic decision model (ISI) as described by Definition 6.1.

Let $\Sigma=(\sigma_0,\ldots,\sigma_{N-1})$ be a statistic sufficient for control with range spaces Y_k, $k=0,\ldots,N-1$. The <u>perfect state information stochastic decision model</u> (PSI) corresponding to (ISI) consists of the following:

$Y_0,\ldots,Y_{N-1}$: State spaces.

C: Control space.

$\hat{\Gamma}_0,\ldots,\hat{\Gamma}_{N-1}$: Constraint spaces defined by (6.6).

α: Discount factor.

$\hat{g}$: One stage cost function defined by (6.5).

$\hat{t}_0,\ldots,\hat{t}_{N-2}$: State transition kernels defined by (6.4).

N: Horizon.

Thus defined, (PSI) is a nonstationary version of (SM) as considered in Chapters 3 - 5. The definitions of policies and cost functions in (PSI) are analogous to those of Chapter 3, Section 1, for (SM). We will use the circumflex (^) to denote these objects in (PSI). For example, $\hat{\Pi}'$ is the set of all policies and $\hat{\Pi}$ is the set of Markov policies in (PSI).

If $\hat{\pi}=(\hat{\mu}_0,\ldots,\hat{\mu}_{N-1})$ is a policy in (PSI), then the sequence

$$(\hat{\mu}_0[du_0|\sigma_0(p;i_0)],\ldots,\hat{\mu}_{N-1}[du_{N-1}|\sigma_0(p;i_0),u_0,\ldots,u_{N-2},\sigma_{N-1}(p;i_{N-1})])$$

is a policy in (ISI). We call this policy $\hat{\pi}$ also and can regard $\hat{\Pi}$ as a subset of Π in this sense. If $\hat{\pi}$ is nonrandomized in (PSI), then it is also nonrandomized in (ISI). We will see in Theorem 6.3 that $\hat{\pi}$ in (PSI) and $\hat{\pi}$ in (ISI) result in the same cost.

Given a policy $\hat{\pi}=(\hat{\mu}_0,\ldots,\hat{\mu}_{N-1})\epsilon\,\hat{\Pi}$ and $q\epsilon P(Y_0)$, there is a sequence of consistent probability measures $\hat{P}_k(\hat{\pi},q)$ generated on $Y_0C\ldots Y_kC$, $k=0,\ldots,N-1$, defined on measurable rectangles by

$$(6.8)\quad \hat{P}_k(\hat{\pi},q)\{\underline{Y}_o\underline{C}_o\ldots\underline{Y}_k\underline{C}_k\} = \int_{\underline{Y}_o}\int_{\underline{C}_o}\cdots\int_{\underline{Y}_k}\hat{\mu}_k(\underline{C}_k|y_o,u_o,\ldots,u_{k-1},y_k)$$

$$\hat{t}_{k-1}(dy_k|y_{k-1},u_{k-1})\ldots\hat{\mu}_o(du_o|y_o)q(dy_o).$$

Define $\varphi:P(S)\to P(Y_o)$ by

$$(6.9)\qquad \varphi(p)(\underline{Y}_o) = \int s_o(\sigma_o(p;\cdot)\varepsilon\underline{Y}_o|x_o)p(dx_o).$$

Thus defined, $\varphi(p)$ is the distribution of the initial state y_o in (PSI) when the initial state x_o in (ISI) has distribution p. The mapping

$$(x_o,p)\to s_o(\sigma_o(p;\cdot)\varepsilon\underline{Y}_o)|x_o) = s_o(\sigma_o^{-1}(\underline{Y}_o)_p|x_o)$$

can be shown to be Borel measurable in the same way the transition kernel t defined by (3.1) was shown to be Borel measurable. Now apply Corollaries B.3.1 and B.3.3 to conclude φ is Borel measurable.

<u>Theorem</u> <u>6.3</u> $(F^+)(F^-)(P)(N)(D)$[3]

If $\hat{\pi}\varepsilon\hat{\Pi}'$ and $p\varepsilon P(S)$, then

$$J_{N,\hat{\pi}}(p) = \int_{Y_o}\hat{J}_{N,\hat{\pi}}(y_o)\varphi(p)(dy_o).$$

Proof:

We show by induction that when $\hat{\pi}\varepsilon\hat{\Pi}'$, $p\varepsilon P(S)$, $k=0,\ldots,N-1$, and $\underline{Y}_o\varepsilon\mathcal{B}_{Y_o},\underline{C}_o\varepsilon\mathcal{B}_C,\ldots,\underline{Y}_k\varepsilon\mathcal{B}_{Y_k},\underline{C}_k\varepsilon\mathcal{B}_C$, then

$$(6.10)\qquad P_k(\hat{\pi},p)\{\sigma_o(p;\cdot)\varepsilon\underline{Y}_o,u_o\varepsilon\underline{C}_o,\ldots,\sigma_k(p;\cdot)\varepsilon\underline{Y}_k,u_k\varepsilon\underline{C}_k\}$$

---

[3]The assumption (P), (N), or (D) is, as defined in Definition 6.3, on the (ISI) model. We have defined the (PSI) model in such a way that if (P), (N), or (D) holds for (ISI), then the same assumption holds for (PSI). The assumption (F^+) or (F^-) is placed on both models, since either one can satisfy such a condition and the other violate it.

$$= \hat{P}(\hat{\pi},\varphi(p))\{\underline{Y}_o\underline{C}_o\ldots\underline{Y}_k\underline{C}_k\}$$

For k=0,

$$P_o(\hat{\pi},p)\{\sigma_o(p;\cdot)\varepsilon\underline{Y}_o,u_o\varepsilon\underline{C}_o\} = \int_{S_o}\int_{\{\sigma_o(p;\cdot)\varepsilon\underline{Y}_o\}}\hat{\mu}_o(\underline{C}_o|\sigma_o(p;z_o))s_o(dz_o|x_o)p(dx_o)$$
by (6.1)
$$= \int_{\underline{Y}_o}\hat{\mu}_o(\underline{C}_o|y_o)\varphi(p)(dy_o) \text{ by (6.9)}$$
$$= \hat{P}_o(\hat{\pi},\varphi(p))\{\underline{Y}_o\underline{C}_o\} \text{ by (6.8).}$$

Assume (6.10) holds for k. Then

$$P_{k+1}(\hat{\pi},p)\{\sigma_o(p;\cdot)\varepsilon\underline{Y}_o,u_o\varepsilon\underline{C}_o,\ldots,\sigma_{k+1}(p;\cdot)\varepsilon\underline{Y}_{k+1},u_{k+1}\varepsilon\underline{C}_{k+1}\}$$

$$= \int_{\{\sigma_o(p;\cdot)\varepsilon\underline{Y}_o,u_o\varepsilon\underline{C}_o,\ldots,\sigma_k(p;\cdot)\varepsilon\underline{Y}_k,u_k\varepsilon\underline{C}_k\}}\int_{\underline{Y}_{k+1}}\hat{\mu}_{k+1}(\underline{C}_{k+1}|\sigma_o(p;i_o),u_o,\ldots$$
$$\ldots,u_k,y_{k+1})\hat{t}_k(dy_{k+1}|\sigma_k(p;i_k),u_k)P_k(\hat{\pi},p) \text{ by (6.1) and (6.4)}$$

$$= \int_{\underline{Y}_o\underline{C}_o\ldots\underline{Y}_k\underline{C}_k}\int_{\underline{Y}_{k+1}}\hat{\mu}_k(\underline{C}_{k+1}|y_o,u_o,\ldots,u_k,y_{k+1})\hat{t}_k(dy_{k+1}|y_k,u_k)d\hat{P}_k(\hat{\pi},\varphi(p))$$
by the induction hypothesis

$$= \hat{P}_{k+1}(\hat{\pi},\varphi(p))\{\underline{Y}_o\underline{C}_o\ldots\underline{Y}_{k+1}\underline{C}_{k+1}\} \text{ by (6.8).}$$

Now

$$J_{N,\hat{\pi}}(p) = \sum_{k=0}^{N-1}\alpha^k\int_{H_k}g(x_k,u_k)dP_k(\hat{\pi},p) \text{ by (6.2)}$$
$$= \sum_{k=0}^{N-1}\alpha^k\int_{H_k}\hat{g}_k(\sigma_k(p;i_k),u_k)dP_k(\hat{\pi},p) \text{ by (6.5)}$$
$$= \sum_{k=0}^{N-1}\alpha^k\int_{Y_oC\ldots Y_kC}\hat{g}_k(y_k,u_k)d\hat{P}_k(\hat{\pi},\varphi(p)) \text{ by (6.10)}$$
$$= \int_{Y_o}[\sum_{k=0}^{N-1}\alpha^k\int_{Y_oC\ldots Y_kC}\hat{g}_k(y_k,u_k)d\hat{P}_k(\hat{\pi},p_{y_o})]\varphi(p)(dy_o)$$
by (6.8) (and the monotone or bounded convergence theorem under (P), (N) or (D))

$$= \int_{Y_o} \hat{J}_{N,\hat{\pi}}(y_o)\varphi(p)(dy_o) \text{ by (3.4) (cf. (3.3) and (6.8)).}$$

QED

<u>Corollary 6.3.1</u> $(F^+)(F^-)(P)(N)(D)$

For every $p\varepsilon P(S)$,

$$J_N^*(p) \leq \int_{Y_o} \hat{J}_N^*(y_o)\varphi(p)(dy_o).$$

Proof:

The function $\hat{J}_N^*$ is lower semianalytic (Corollary 3.2.1 or Corollary 4.5.1), so the above integral is defined. For $p\varepsilon P(S)$,

$$J_N^*(p) = \inf_{\pi\varepsilon\Pi} J_{N,\pi}(p) \leq \inf_{\hat{\pi}\varepsilon\hat{\Pi}'} J_{N,\hat{\pi}}(p)$$

$$= \inf_{\hat{\pi}\varepsilon\hat{\Pi}'} \int_{Y_o} \hat{J}_{N,\hat{\pi}}(y_o)\varphi(p)(dy_o) \text{ by Theorem 6.3}$$

$$= \int_{Y_o} \hat{J}_N^*(y_o)\varphi(p)(dy_o) \text{ by Corollary 3.2.4 or Corollary 4.6.2.}$$

QED

We now show that J_N^* and $\hat{J}_N^*$ correspond in the same way that $J_{N,\hat{\pi}}$ and $\hat{J}_{N,\hat{\pi}}$ correspond in Theorem 6.3.

<u>Lemma 6.1</u> $(F^+)(F^-)(P)(N)(D)$

For $\pi\varepsilon\Pi$, $p\varepsilon P(S)$, there exists $\hat{\pi}\varepsilon\hat{\Pi}$ for which

$$J_{N,\pi}(p) = \int_{Y_o} \hat{J}_{N,\hat{\pi}}(y_o)\varphi(p)(dy_o).$$

Proof:

Let $\pi=(\mu_o,\ldots,\mu_{N-1})\varepsilon\Pi$ and $p\varepsilon P(S)$. Let $Q_k(\pi,p)$ be probability measures on Y_kC_k, $k=0,\ldots,N-1$, defined on measurable rectangles by

$$\text{(6.11)} \qquad Q_k(\pi,p)\{\underline{Y}_k\underline{C}_k\} = P_k(\pi,p)\{\sigma_k(p;\cdot)\varepsilon\underline{Y}_k, u_k\varepsilon\underline{C}_k\}.$$

By Theorem C.2, there exist Borel measurable stochastic kernels $\hat{\mu}_k(du_k|y_k)$ on C given Y_k satisfying

$$(6.12) \qquad Q_k(\pi,p)(\underline{Y}_k\underline{C}_k) = \int_{\underline{Y}_k} \hat{\mu}_k(\underline{C}_k|y_k)Q_k(\pi,p)(dy_k \cdot C),$$

for every $\underline{Y}_k \varepsilon \mathcal{B}_Y$, $\underline{C}_k \varepsilon \mathcal{B}_C$, $k=0,\ldots,N-1$.

Then for $k=0,\ldots,N-1$,

$$\begin{aligned} 1 &= P_k(\pi,p)\{(i_k,u_k)\varepsilon\Gamma_k\} \\ &= P_k(\pi,p)\{(\sigma_k(p;\cdot),u_k)\varepsilon\hat{\Gamma}_k\} \text{ by (6.7)} \\ &= Q_k(\pi,p)(\hat{\Gamma}_k) \\ &= \int_{Y_k} \hat{\mu}_k((\hat{\Gamma}_k)_{y_k}|y_k)Q_k(\pi,p)(dy_k \cdot C), \end{aligned}$$

which implies $\hat{\mu}_k((\hat{\Gamma})_{y_k}|y_k)=1$ $Q_k(\pi,p)$-almost surely. Redefining $\hat{\mu}_k(\cdot|y_k)$ on a set of $Q_k(\pi,p)$ measure zero if necessary, we can assume that (6.12) holds and $\hat{\pi}=(\hat{\mu}_0,\ldots,\hat{\mu}_{N-1})\varepsilon\,\ddot{\Pi}$.

We show by induction that for $\underline{Y}_k \varepsilon \mathcal{B}_Y$, $\underline{C}_k \varepsilon \mathcal{B}_C$, and $k=0,\ldots,N-1$,

$$(6.13) \qquad Q_k(\pi,p)(\underline{Y}_k\underline{C}_k) = \hat{P}_k(\hat{\pi},\varphi(p))\{y_k\varepsilon\underline{Y}_k, u_k\varepsilon\underline{C}_k\}.$$

For $k=0$,

$$\begin{aligned} Q_0(\pi,p)(\underline{Y}_0\underline{C}_0) &= \int_{\underline{Y}_0} \hat{\mu}_0(\underline{C}_0|y_0)\varphi(p)(dy_0) \text{ by (6.1), (6.9), (6.11) and (6.12)} \\ &= \hat{P}_0(\hat{\pi},\varphi(p))\{y_0\varepsilon\underline{Y}_0, u_0\varepsilon\underline{C}_0\} \text{ by (6.8).} \end{aligned}$$

Assume (6.13) holds for k. Then

$$\begin{aligned} Q_{k+1}(\pi,p)(\underline{Y}_{k+1}\underline{C}_{k+1}) &= \int_{\underline{Y}_{k+1}} \hat{\mu}_{k+1}(\underline{C}_{k+1}|y_{k+1})Q_{k+1}(\pi,p)(dy_{k+1} \cdot C) \text{ by (6.12)} \\ &= \int_{\{\sigma_{k+1}(p;\cdot)\varepsilon\underline{Y}_{k+1}\}} \hat{\mu}_{k+1}(\underline{C}_{k+1}|\sigma_{k+1}(p;i_{k+1}))dP_{k+1}(\pi,p) \\ &\quad \text{by (6.11)} \end{aligned}$$

$$= \int_{H_k}\int_{\underline{Y}_{k+1}} \hat{\mu}_{k+1}(\underline{C}_{k+1}|y_{k+1})\hat{t}_k(dy_{k+1}|\sigma_k(p;i_k),u_k)dP_k(\pi,p)$$

by (6.4) and the consistency of the $P_k(\pi,p)$'s

$$= \int_{Y_k C}\int_{\underline{Y}_{k+1}} \hat{\mu}_{k+1}(\underline{C}_{k+1}|y_{k+1})\hat{t}_k(dy_{k+1}|y_k,u_k)dQ_k(\pi,p)$$

by (6.11)

$$= \int_{Y_o C\ldots Y_k C}\int_{\underline{Y}_{k+1}} \hat{\mu}_{k+1}(\underline{C}_{k+1}|y_{k+1})\hat{t}(dy_{k+1}|y_k,u_k)d\hat{P}_k(\hat{\pi},\varphi(p))$$

by the induction hypothesis

$$= \hat{P}_{k+1}(\hat{\pi},\varphi(p))\{y_{k+1}\varepsilon\underline{Y}_{k+1},u_{k+1}\varepsilon\underline{C}_{k+1}\} \text{ by (6.8).}$$

Finally,

$$J_{N,\pi}(p) = \sum_{k=0}^{N-1}\alpha^k\int g(x_k,u_k)dP_k(\pi,p)$$

$$= \sum_{k=0}^{N-1}\alpha^k\int \hat{g}_k(y_k,u_k)d\hat{P}_k(\hat{\pi},\varphi(p)) \text{ by (6.5), (6.11) and (6.13)}$$

$$= \int_{Y_o}[\sum_{k=0}^{N-1}\alpha^k\int \hat{g}_k(y_k,u_k)d\hat{P}_k(\hat{\pi},p_{y_o})]\varphi(p)(dy_o)$$

by (6.8) (and the monotone or bounded convergence theorem under (P), (N) or (D))

$$= \int_{Y_o}\hat{J}_{N,\hat{\pi}}(y_o)\varphi(p)(dy_o) \text{ by (3.4) (cf. (3.3) and (6.8)).} \quad \text{QED}$$

<u>Theorem</u> <u>6.4</u> $(F^+)(F^-)(P)(N)(D)$[4]

For $p\varepsilon P(S)$,

$$(6.14) \qquad J_N^*(p) = \int_{Y_o}\hat{J}_N^*(y_o)\varphi(p)(dy_o).$$

Also, if $\hat{\pi}$ is optimal, $\varphi(p)$-optimal or weakly $\varphi(p)$-ϵ-optimal in (PSI), then $\hat{\pi}$ is optimal, optimal at p or ϵ-optimal at p, respectively, in (ISI).

[4]See Footnote 3.

Proof:

Equation (6.14) follows from Corollary 6.3.1 and Lemma 6.1. If $\hat{\pi}$ is weakly $\varphi(p)$-ϵ-optimal in (PSI), then by Theorem 6.3

$$\begin{aligned} J_{N,\hat{\pi}}(p) &= \int_{Y_0} \hat{J}_{N,\hat{\pi}}(y_0)\varphi(p)(dy_0) \\ &\leq \int_{Y_0} \hat{J}_N^*(y_0)\varphi(p)(dy_0) + \epsilon \text{ if} \int \hat{J}_N^*(y_0)\varphi(p)(dy_0) > -\infty; \\ &\leq -1/\epsilon \text{ if} \int \hat{J}_N^*(y_0)\varphi(p)(dy_0) = -\infty. \end{aligned}$$

Equation (6.14) implies that $\hat{\pi}$ is ϵ-optimal at p. If $\hat{\pi}$ is optimal or $\varphi(p)$-optimal in (PSI), a similar argument using Theorem 6.3 and (6.14) shows that $\hat{\pi}$ is optimal or optimal at p, respectively, in (ISI).

We will show shortly that a statistic sufficient for control always exists and, indeed, in many cases can be chosen so that (PSI) is stationary. The existence of such a statistic for (ISI) and the consequent existence of the corresponding (PSI) enable us to utilize the results of Chapters 3, 4 and 5. For example, we have the following corollary to Theorem 6.4.

<u>Corollary 6.4.1</u> $(F^+)(F^-)(P)(N)(D)$

If a statistic $\Sigma = (\sigma_0, \ldots, \sigma_{N-1})$ sufficient for control exists for (ISI), then for every $\epsilon > 0$, there exists an ϵ-optimal nonrandomized policy for (ISI) which depends on i_k only through $\sigma_k(p;i_k)$, i.e. has the form

$$\pi = (\mu_0(p;\sigma_0(p;i_0)), \ldots, \mu_{N-1}(p;\sigma_{N-1}(p;i_{N-1}))).$$

Proof:

Apply Theorems 6.2 and 6.4. QED

The other specific results which can be derived for (ISI) from Chapters 3 - 5 are obvious and will not be exhaustively listed. We content ourselves with describing the dynamic programming algorithm over a finite horizon.

By Theorem 3.2, the dynamic programming algorithm has the following form under (F^+) and (F^-), where we assume (PSI) is stationary:

(6.15) $\hat{J}_0^*(y) = 0$ for every y,

(6.16) $\hat{J}_{k+1}^*(y) = \inf_{u \in \hat{\Gamma}_y} [\hat{g}(y,u) + \alpha \int \hat{J}_k^*(y')\hat{t}(dy'|y,u)]$, $k=0,\ldots,N-1$.

If the infimum in (6.16) is achieved for every y and $k=0,\ldots,N-1$, then there exist universally measurable functions $\hat{\mu}_k : Y \rightarrow C$ such that for every y and $k=0,\ldots,N-1$, $\hat{\mu}_k(y) \in \hat{\Gamma}_y$ and $\hat{\mu}_k(y)$ achieves the infimum in (6.16). Then $\hat{\pi}=(\hat{\mu}_0,\ldots,\hat{\mu}_{N-1})$ is optimal in (PSI) (Theorem 3.3) and $(\hat{\mu}_0 \circ \sigma_0,\ldots,\hat{\mu}_{N-1} \circ \sigma_{N-1})$ is optimal in (ISI) (Theorem 6.4).

In many cases $\sigma_{k+1}(p;i_{k+1})$ is a function of $\sigma_k(p;i_k)$, u_k and z_{k+1}. The computational procedure in such a case is to first construct $(\hat{\mu}_0,\ldots,\hat{\mu}_{N-1})$ via (6.15) and (6.16), then compute $y_0=\sigma_0(p;i_0)$ from the initial distribution and the initial observation, and apply control $u_0=\hat{\mu}_0(y_0)$. Given y_k, u_k and z_{k+1}, compute y_{k+1} and apply control $u_{k+1}=\hat{\mu}_{k+1}(y_{k+1})$, $k=0,\ldots,N-2$. In this way the information contained in $(p;i_k)$ has been condensed into y_k. This condensation of information is the historical motivation for statistics sufficient for control, but is peripheral to the theoretical development here.

Turning to the question of the existence of a statistic sufficient for control, it is not surprising to discover that the sequence of identity mappings on $P(S)I_k$, $k=0,\ldots,N-1$, is such an object. Although this represents no condensation of information, it is sufficient to justify our analysis thus far. If the constraint sets Γ_k are equal to I_kC, $k=0,\ldots,N-1$, then the

functions mapping $P(S)I_k$ into the distribution of x_k conditioned on $(p;i_k)$, $k=0,\ldots,N-1$, constitute a statistic sufficient for control. This statistic has the property that its value at the $(k+1)$-st stage is a function of its value at the k-th stage, u_k and z_{k+1} (cf. (6.21)), so it represents a genuine condensation of information. It also results in a stationary perfect state information model and, if these distributions can be characterized by a set of parameters, results in enormous computational simplification. This latter condition is the case, for example, if it is possible to show before-hand that all these distributions are Gaussian.

We prove these facts in reverse order.

Lemma 6.2 There exist Borel measurable stochastic kernels $r_o(dx_o|p;z_o)$ on S given $P(S)Z$ and $r(dx|p;u,z)$ on S given $P(S)CZ$ which satisfy

$$(6.17)\qquad \int_{\underline{S}_o} s_o(\underline{Z}_o|x_o)p(dx_o) = \int_S\int_{\underline{Z}_o} r_o(\underline{S}_o|p;z_o)s_o(dz_o|x_o)p(dx_o)$$

for every $\underline{S}_o\in\mathcal{B}_S$, $\underline{Z}_o\in\mathcal{B}_Z$, $p\in P(S)$, and

$$(6.18)\qquad \int_{\underline{S}} s(\underline{Z}|u,x)p(dx) = \int_S\int_{\underline{Z}} r(\underline{S}|p;u,z)s(dz|u,x)p(dx)$$

for every $\underline{S}\in\mathcal{B}_S$, $\underline{Z}\in\mathcal{B}_Z$, $p\in P(S)$ and $u\in C$.

Proof:

For fixed $(p;u)$, define a measure q on SZ by specifying its values on measurable rectangles to be

$$q(\underline{SZ}|p;u) = \int_{\underline{S}} s(\underline{Z}|u,x)p(dx).$$

Then q is a Borel measurable stochastic kernel on SZ given $P(S)C$ (Corollary B.3.1) and can be decomposed into $r(dx|p;u,z)$ and $q(S\cdot dz|p;u)$ (Theorem C.2). Equation (6.18) follows and (6.17) is a special case of (6.18). QED

It is customary to call p, the given distribution of x_o, the a priori distribution of the initial state. After z_o is observed, the distribution is "up-dated", i.e. the distribution of x_o conditioned on z_o is computed. The up-dated distribution is called the a posteriori distribution and is just $r_o(\cdot|p;z_o)$. At the k-th stage, $k\geq 1$, we will have some a priori distribution p'_k of x_k based on $i_{k-1}=(z_o,u_o,\ldots,u_{k-2},z_{k-1})$. Control u_{k-1} is applied, some z_k is observed, and an a posteriori distribution of x_k conditioned on (i_{k-1},u_{k-1},z_k) is computed. This distribution is just $r(\cdot|p'_k;u_{k-1},z_k)$. The process of passing from an a priori to an a posteriori distribution in this manner is called filtering. Theorem C.2 is crucial in establishing the filtering equations (6.17) and (6.18).

Define $\bar{f}_u:P(S)\rightarrow P(S)$ by

$$\bar{f}_u(p)(\underline{S}) = \int t(\underline{S}|x,u)p(dx),\quad \underline{S}\in\mathcal{B}_S. \tag{6.19}$$

Equation (6.19) can be termed the one-stage prediction equation. If x_k has a posteriori distribution p_k and the control u_k is chosen, then the a priori distribution of x_{k+1} is $\bar{f}_{u_k}(p_k)$. This will be up-dated to the a posteriori distribution as soon as z_{k+1} is observed (cf. (6.21))

The mapping $(u,p)\rightarrow\bar{f}_u(p)$ is Borel measurable (Corollaries B.3.1 and B.3.3). Given a sequence $i_k\in I_k$ such that $i_{k+1}=(i_k,u_k,z_{k+1})$, k=0,...,N-2, and given $p\in P(S)$, define recursively

$$p_o(p;i_o) = r_o(\cdot|p;z_o), \tag{6.20}$$

$$p_{k+1}(p;i_{k+1}) = r(\cdot|\bar{f}_{u_k}[p_k(p;i_k)];u_k,z_{k+1}),\quad k=0,\ldots,N-2. \tag{6.21}$$

Lemma 6.3 Let $p\in P(S)$ and $\pi=(\mu_o,\ldots,\mu_{N-1})\in\Pi$ be given. Then for $S_k\in\mathcal{B}_S$,

(6.22) $$P_k(\pi,p)\{x_k \varepsilon \underline{S}_k \mid i_k\} = p_k(p;i_k)(\underline{S}_k)$$

$P_k(\pi,p)$-almost surely, $k=0,\ldots,N-1$.

Proof:

We proceed by induction. By definition

$$\int_{\{z_o \varepsilon \underline{Z}_o\}} p_o(p;i_o)(\underline{S}_o)dP_o(\pi,p) = \int_{\{z_o \varepsilon \underline{Z}_o\}} r_o(\underline{S}_o \mid p;z_o)dP_o(\pi,p)$$

$$= \int_S \int_{\underline{Z}_o} r(\underline{S}_o \mid p;z_o)s_o(dz_o \mid x_o)p(dx_o)$$

$$= \int_{\underline{S}_o} s_o(\underline{Z}_o \mid x_o)p(dx_o) \text{ by (6.17)}$$

$$= P_o(\pi,p)\{x_o \varepsilon \underline{S}_o, z_o \varepsilon \underline{Z}_o\}$$

for $\underline{S}_o \varepsilon \mathcal{B}_S$, $\underline{Z}_o \varepsilon \mathcal{B}_Z$, and so by the definition of conditional probability

$$P_o(\pi,p)\{x_o \varepsilon \underline{S}_o \mid i_o\} = p_o(p;i_o)(\underline{S}_o)$$

$P_o(\pi,p)$-almost surely.

Assume that (6.22) holds for k. For $\underline{I}_k \varepsilon \mathcal{B}_{I_k}$, $\underline{C}_k \varepsilon \mathcal{B}_C$, $\underline{Z}_{k+1} \varepsilon \mathcal{B}_Z$ and $\underline{S}_{k+1} \varepsilon \mathcal{B}_S$,

$$\int_{\{i_k \varepsilon \underline{I}_k, u_k \varepsilon \underline{C}_k, z_{k+1} \varepsilon \underline{Z}_{k+1}\}} p_{k+1}(p;i_k,u_k,z_{k+1})(\underline{S}_{k+1})dP_{k+1}(\pi,p)$$

$$= \int_{\{i_k \varepsilon \underline{I}_k\}} \int_{\underline{C}_k} \int_{S_{k+1}} \int_{\underline{Z}_{k+1}} p_{k+1}(p;i_k,u_k,z_{k+1})(\underline{S}_{k+1})s(dz_{k+1} \mid u_k,x_{k+1})$$

$$t(dx_{k+1} \mid x_k,u_k)\mu_k(du_k \mid p;i_k)dP_k(\pi,p) \text{ by (6.1)}$$

$$= \int_{\{i_k \varepsilon \underline{I}_k\}} \int_{S_k} \int_{\underline{C}_k} \int_{S_{k+1}} \int_{\underline{Z}_{k+1}} p_{k+1}(p;i_k,u_k,z_{k+1})(\underline{S}_{k+1})s(dz_{k+1} \mid u_k,x_{k+1})$$

$$t(dx_{k+1} \mid x_k,u_k)\mu_k(du_k \mid p;i_k)[p_k(p;i_k)(dx_k)]dP_k(\pi,p)$$

by the induction hypothesis

$$= \int_{\{i_k \varepsilon \underline{I}_k\}} \int_{\underline{C}_k} \int_{S_k} \int_{S_{k+1}} \int_{\underline{Z}_{k+1}} p_{k+1}(p;i_k,u_k,z_{k+1})(\underline{S}_{k+1}) s(dz_{k+1} \mid u_k,x_{k+1})$$

$t(dx_{k+1} \mid x_k,u_k)[p_k(p;i_k)(dx_k)]\mu_k(du_k \mid p;i_k)dP_k(\pi,p)$ by Fubini's Theorem

$$= \int_{\{i_k \varepsilon \underline{I}_k\}} \int_{\underline{C}_k} \int_{S_{k+1}} \int_{\underline{Z}_{k+1}} r(\underline{S}_{k+1} \mid \bar{f}_{u_k}[p_k(p;i_k)];u_k,z_{k+1}) s(dz_{k+1} \mid u_k,x_{k+1})$$

$\bar{f}_{u_k}[p_k(p;i_k)](dx_{k+1})\mu_k(du_k \mid p;i_k)dP_k(\pi,p)$ by (6.19) and (6.21)

$$= \int_{\{i_k \varepsilon \underline{I}_k\}} \int_{\underline{C}_k} \int_{\underline{S}_{k+1}} s(\underline{Z}_{k+1} \mid u_k,x_{k+1})\bar{f}_{u_k}[p_k(p;i_k)](dx_{k+1})$$

$\mu_k(du_k \mid p;i_k)dP_k(\pi,p)$ by (6.18)

$$= \int_{\{i_k \varepsilon \underline{I}_k\}} \int_{\underline{C}_k} \int_{S_k} \int_{\underline{S}_{k+1}} s(\underline{Z}_{k+1} \mid u_k,x_{k+1})t(dx_{k+1} \mid x_k,u_k)[p_k(p;i_k)(dx_k)]$$

$\mu_k(du_k \mid p;i_k)dP_k(\pi,p)$ by (6.19)

$$= \int_{\{i_k \varepsilon \underline{I}_k\}} \int_{S_k} \int_{\underline{C}_k} \int_{\underline{S}_{k+1}} s(\underline{Z}_{k+1} \mid u_k,x_{k+1})t(dx_{k+1} \mid x_k,u_k)\mu_k(du_k \mid p;i_k)$$

$[p_k(p;i_k)(dx_k)]dP_k(\pi,p)$ by Fubini's Theorem

$$= \int_{\{i_k \varepsilon \underline{I}_k\}} \int_{\underline{C}_k} \int_{\underline{S}_{k+1}} s(\underline{Z}_{k+1} \mid u_k,x_{k+1})t(dx_{k+1} \mid x_k,u_k)\mu_k(du_k \mid p;i_k)$$

$dP_k(\pi,p)$ by the induction hypothesis

$$= P_{k+1}(\pi,p)\{i_k \varepsilon \underline{I}_k, u_k \varepsilon \underline{C}_k, x_{k+1} \varepsilon \underline{S}_{k+1}, z_{k+1} \varepsilon \underline{Z}_{k+1}\}.$$

Therefore (6.22) holds for k+1. QED

Theorem 6.5 If $\Gamma_k = I_k C$, $k=0,\ldots,N-1$, then $\{p_k(p;i_k)\}$ as defined by (6.20) and (6.21) is a statistic sufficient for control and the resulting perfect state information model is stationary.

Proof:

Let Y_k in Definition 6.4 be $P(S)$, $k=0,\ldots,N-1$. The mapping p_o of (6.20) is Borel measurable, and p_{k+1} is the composition of Borel measurable mappings

whenever p_k is Borel measurable. Thus $(p_0,\ldots,p_{N-1})$ is a statistic.

Condition (a) of Definition 6.4 is satisfied when $\Gamma_k = I_k C$, $k=0,\ldots,N-1$. We verify (6.4) and (6.5).

For $y \in P(S)$, $u \in C$ and $\underline{Y} \in \mathcal{B}_{P(S)}$, define

$$\underline{Z}(y,u,\underline{Y}) = \{z \in Z\colon\ r(\cdot|\bar{f}_u(y);u,z) \in \underline{Y}\},$$

$$\hat{t}(\underline{Y}|y,u) = \int_S \int_S s(\underline{Z}(y,u,\underline{Y})|u,x')t(dx'|x,u)y(dx).$$

Note that $\underline{Z}(y,u,\underline{Y})$ is the (y,u)-section of the inverse image of $\underline{Y}$ under a Borel measurable mapping, and the measure on Z defined by

$$\int_S \int_S s(\underline{Z}|u,x')t(dx'|x,u)y(dx),\ \underline{Z} \in \mathcal{B}_Z,$$

depends Borel measurably on (y,u) (Corollaries B.3.1 and B.3.3). Consequently, $\hat{t}$ can be shown to be a Borel measurable stochastic kernel on $P(S)$ given $P(S)C$ by the same argument as was used to show the state transition kernel t (Definition 3.2) is a Borel measurable stochastic kernel.

We have for $\pi \in \Pi$, $p \in P(S)$, and $k=0,\ldots,N-2$,

$$P_{k+1}(\pi,p)\{p_{k+1}(p;\cdot) \in \underline{Y} | p_k(p;\cdot)=y, u_k=u\}$$

$$= P_{k+1}(\pi,p)\{z_{k+1} \in \underline{Z}(y,u,\underline{Y}) | p_k(p;\cdot)=y, u_k=u\}$$

$$= E\{E\{\chi_{\{z_{k+1} \in \underline{Z}(y,u,\underline{Y})\}} | i_k, u_k\} | p_k(p;\cdot)=y, u_k=u\}$$

$$= E\{\int_S \int_S s(\underline{Z}(y,u,\underline{Y})|u_k,x_{k+1})t(dx_{k+1}|x_k,u_k)[p_k(p;i_k)(dx_k)] | p_k(p;\cdot)=y, u_k=u\}$$

by Lemma 6.3

$$= \hat{t}(\underline{Y}|y,u)$$

$P_{k+1}(\pi,p)$-almost surely, where the expectations are with respect to $P_{k+1}(\pi,p)$. Thus (6.4) is satisfied. Note that $\hat{t}$ is independent of k.

For $k=0,\ldots,N-1$, $\pi\varepsilon\Pi$ and $p\varepsilon P(S)$, by Lemma 6.3

$$E\{g(x_k,u_k)|p_k(p;\cdot)=y,u_k=u\} = \int_S g(x_k,u)y(dx_k) = \hat{g}(y,u)$$

$P_{k+1}(\pi,p)$-almost surely, where the expectation is with respect to $P_{k+1}(\pi,p)$. Note that $\hat{g}$ is lower semianalytic (Theorem 2.4) and independent of k.

<u>Theorem</u> <u>6.6</u> The set of identity mappings on $P(S)I_k$, $k=0,\ldots,N-1$, is a statistic sufficient for control.

Proof:

Let Y_k in Definition 6.4 be $P(S)I_k$ and let σ_k be the identity mapping on $P(S)I_k$, $k=0,\ldots,N-1$. Condition (a) of Definition 6.4 is clearly satisfied. We verify (6.4) and (6.5).

For $\pi\varepsilon\Pi$, $p\varepsilon P(S)$, and $k=0,\ldots,N-2$, the distribution

$$P_{k+1}(\pi,p)\{d\sigma_{k+1}|\sigma_k(p;i_k)=y,u_k=u\}$$

depends Borel measurably on

$$P_{k+1}(\pi,p)\{dz_{k+1}|\sigma_k(p;i_k)=y,u_k=u\}.$$

But for $\underline{Z}\varepsilon\mathcal{B}_Z$, by Lemma 6.3

$$P_{k+1}(\pi,p)\{z_{k+1}\varepsilon\underline{Z}|\sigma_k(p;\cdot)=y,u_k=u\} = \int_S\int_S s(\underline{Z}|x',u)t(dx'|x,u)[p_k(y)(dx)]$$

$P_{k+1}(\pi,p)$-almost surely. This last expression is Borel measurable in (y,u), so (6.4) holds.

For $k=0,\ldots,N-2$, $\pi\varepsilon\Pi$, and $p\varepsilon P(S)$, by Lemma 6.3

$$E\{g(x_k,u_k)|\sigma_k(p;\cdot)=y,u_k=u\} = \int_S g(x_k,u)[p_k(y)(dx_k)] \tag{6.23}$$

$P_{k+1}(\pi,p)$-almost surely and (6.5) holds. The right hand side of (6.23) is lower semianalytic in (y,u) by Theorem 2.4. QED

APPENDIX A

ANALYTIC SETS

This appendix summarizes the pertinent facts about analytic sets available in the literature. There are several equivalent definitions of analytic sets in use. The one given here is a variation of that found on page 15 of [26].

<u>Definition A.1</u> Let N be the cross product of countably many copies of the positive integers. Let the set of positive integers have the discrete topology and N the product topology. A separable metric set A is <u>analytic</u> if there is a continuous function f mapping N onto A.

The definition in [26] requires that A be embedded in a complete separable metric space. Given an A satisfying Definition A.1, it can be embedded in its metric completion without affecting the continuity of f, so this requirement is really superfluous.

Note also that N as defined in Definition A.1 is homeomorphic to N', the set of irrationals in (0,1) with the usual topology [19, p. 25]. N could be replaced by N' in Definition A.1, and indeed this is the characterization of analytic sets found in Section 39 of [19].

Theorems A.1 - A.3 are proved in Chapter I, Section 3, of [26].

<u>Theorem A.1</u> Let X be a Borel space, i.e. a Borel subset of a complete separable topological space. Then X is analytic.

<u>Theorem A.2</u> The countable union, intersection and cross product of analytic sets is analytic.

<u>Theorem A.3</u> Let A and B be analytic subsets of Borel spaces X and Y respectively. If f is a Borel measurable function from X to Y, then f(A) and $f^{-1}(B)$ are analytic.

<u>Corollary A.3.1</u> Let X and Y be Borel spaces and A an analytic subset of the Cartesian product XY. Then

$$\mathrm{proj}_X A = \{x \in X: \text{ for some } y,\ (x,y) \in A\}$$

is analytic.

Proof:

The projection mapping is Borel measurable, in fact continuous, from XY to X. QED

<u>Definition A.2</u> Let $\{A(n_1,\dots,n_k)\}$ be a system of sets in some space, where $(n_1,\dots,n_k)$ ranges over the set of finite sequences of positive integers. The set

$$R = \bigcup_{(n_1,n_2,\dots)} \bigcap_{k=1}^{\infty} A(n_1,\dots,n_k)$$

is called the <u>result of operation (A)</u> applied to the system $\{A(n_1,\dots,n_k)\}$.

<u>Definition A.3</u> The system $\{A(n_1,\dots,n_k)\}$ is <u>regular</u> if

$$A(n_1,\dots,n_k,n_{k+1}) \subset A(n_1,\dots,n_k)$$

for each $(n_1,n_2,\dots)$ and k.

<u>Definition A.4</u> A collection $\mathcal{S}$ of subsets of a space is <u>invariant under the operation (A)</u> if whenever each of the sets of the system $\{A(n_1,\dots,n_k)\}$ is in $\mathcal{S}$, the result of operation (A) applied to the system is in $\mathcal{S}$.

The proof of the next theorem can be found in [30], Chapter II, Section 5.

<u>Theorem A.4</u> Let X be a Borel space and m a measure on $\mathcal{B}_X$, the σ-algebra of Borel subsets of X. Let $\mathcal{S}$ be the completion of $\mathcal{B}_X$ with respect to m. Then $\mathcal{S}$ is invariant under operation (A).

<u>Theorem A.5</u> Let X be a complete separable metric space and let $\{A(n_1,\ldots,n_k)\}$ be a regular system of closed subsets of X such that for each fixed $(n_1, n_2,\ldots)$, $\text{diameter}(A(n_1,\ldots,n_k)) \to 0$ as $k \to \infty$. Then the result of performing operation (A) on this system is an analytic set. Conversely, every analytic subset of X can be obtained in this manner.

See [26], Chapter I, Section 3, for a proof.

<u>Definition A.5</u> Given a Borel space X, the intersection of all completions with respect to finite measures of the Borel σ-algebra $\mathcal{B}_X$ is called the <u>universal σ-algebra</u> $\mathcal{U}_X$. A set in $\mathcal{U}_X$ is said to be <u>universally measurable</u>.

If X is a Borel space and m is a finite measure on $\mathcal{B}_X$, then m has a unique extension to a measure on $\mathcal{U}_X$. We denote this extension by m also, writing m(E) instead of m*(E) when $E \in \mathcal{U}_X$. Likewise, if f is a real-valued function on X, measurable with respect to $\mathcal{U}_X$, we will write $\int f\,dm$ to indicate the integral of f with respect to the completion of m on $\mathcal{B}_X$.

<u>Corollary A.5.1</u> Let X be a Borel space. Every analytic subset of X is universally measurable.

Proof:

Given any finite measure m on $\mathcal{B}_X$, let $\mathcal{S}(m)$ denote the completion of $\mathcal{B}_X$ with respect to m. Let A be an analytic subset of X. By Theorem A.5, A is

the result of operation (A) applied to a system of sets in $\mathcal{S}(m)$. By Theorem A.4, A is in $\mathcal{S}(m)$. QED

Definition A.6 Given a Borel space X, the analytic σ-algebra $\mathcal{A}_X$ is the smallest σ-algebra in X containing the class of analytic sets.

The analytic σ-algebra is strictly larger than the class of analytic sets, since the complement of an analytic set is analytic only in the special case that both are Borel [26, p. 20]. Analytic sets which are not Borel do exist [19, p. 460].

We have the following selection theorem originally proved by von Neumann [22]. The version given here can be found in [6].

Theorem A.6 Let X and Y be Borel sets and A an analytic subset of XY. Then there exists a function $\varphi: \mathrm{proj}_X A \rightarrow Y$ such that $(x, \varphi(x)) \in A$ for every $x \in \mathrm{proj}_X A$ and $\varphi^{-1}(B) \in \mathcal{A}_X$ for every Borel subset B of X.[1]

Theorem A.7 The analytic subsets of a Borel space X coincide with the projections on the X-axis of the closed sets in XN', where N' is the set of irrationals in (0,1) with the usual topology.

Proof:

By Theorem A.3, the projections on the X-axis of the closed sets in XN' are analytic. Now let A be an analytic subset of X. Then A = f(N'), where f is continuous, and so the mapping $(x,z) \rightarrow d(x,f(z))$ is continuous from XN' to R, where d is a metric on X consistent with its topology. The inverse image of {0} under this mapping, which is $\{(x,z): x=f(z)\}$, is closed, and A is the projection of this set on the X-axis. QED

[1]This latter property is called analytic measurability (Definition 2.2).

APPENDIX B

MEASURABLE SETS OF MEASURES

This appendix collects standard results about the space of probability measures on a Borel space. Some variations of these results are proved.

Given a Borel space X, i.e. a Borel subset of a complete separable topological space, denote by C(X) the set of bounded continuous real-valued functions on X and by P(X) the set of probability measures on $(X, \mathcal{B}_X)$. If d is a metric on X consistent with the given topology, then $U_d(X)$ will denote the set of functions in C(X) which are uniformly continuous with respect to d.

Following [26], Chapter II, Section 6, define a topology on P(X) by taking the basic open sets to be the class $\mathcal{U}_1$ of sets of the form

$$V_p(f_1,\ldots,f_k;\epsilon_1,\ldots,\epsilon_k) = \{q \in P(X) : |\int f_i dq - \int f_i dp| < \epsilon_i,\ i=1,\ldots,k\},$$

where $f_i \in C(X)$, $\epsilon_i > 0$, $i=1,\ldots,k$; $p \in P(X)$; and k is a positive integer. We will always understand P(X) to be equipped with this topology and will denote by $\mathcal{B}_{P(X)}$ the σ-algebra generated by the open sets in P(X). This notation will be justified by Corollary B.6.1, which states that P(X) is also a Borel space. It follows from the metrizability and separability of X that P(X) is metrizable and separable [26, Chapter II].

<u>Theorem B.1</u> Let $\{p_n\}$ be a sequence in P(X). The following statements are equivalent:

(a) $p_n \to p$;

(b) for every f in C(X), $\lim_n \int f\, dp_n = \int f\, dp$;

(c) for some metric d consistent with the topology on X and every g in $U_d(X)$, $\lim_n \int g\, dp_n = \int g\, dp$;

(d) $\lim\sup_n p_n(F) \leq p(F)$ for every closed set F;

(e) $\lim\inf_n p_n(G) \geq p(G)$ for every open set G;

(f) $\lim_n p_n(B) = p(B)$ for every Borel set B whose boundary has p measure zero.

Proof:

The equivalence of (a) and (b) is a simple consequence of the definition of the topology on P(X). Clearly (b) implies (c). If (c) holds, then since the topology on P(X) depends only on the topology of X and not on any particular metrization, (a) must hold by [26], Chapter II, Theorem 6.1 and the metrizability of P(X). The other equivalences follow from the same theorem. QED

We now exhibit a countable basis which generates the topology in P(X).

<u>Theorem B.2</u> Let X be a Borel space. Let D be dense in P(X). Then there exists a sequence $\{g_1,g_2,\ldots\}$ in C(X) such that the topology on P(X) is generated by the class of basic open sets

$$\mathcal{U}_3 = \{V_p(g_1,\ldots,g_k;\ \epsilon_1,\ldots,\epsilon_k):\ p\in D,\ \epsilon_i>0 \text{ rational, } k \text{ a positive integer}\}.$$

Proof:

By [26], Chapter II, Theorem 6.6, there is a sequence $\{g_1,g_2,\ldots\}$ of functions in C(X) such that whenever $\{p_n\}$ is a sequence in P(X), $p_n \to p$ if and only if $\int g_k dp_n \to \int g_k dp$ for every k.

Let $\mathcal{T}_1$ be the standard topology on P(X), i.e. the topology generated by $\mathcal{U}_1$. Define

$$\mathcal{U}_2 = \{V_p(g_1,\ldots,g_k;\ \epsilon_1,\ldots,\epsilon_k):\ \epsilon_i>0,\ k \text{ a positive integer}\},$$

and take $\mathfrak{J}_2$ to be the topology for which $\mathcal{U}_2$ is a basis. Denote by $\mathfrak{J}_3$ the topology generated by $\mathcal{U}_3$. Then $\mathfrak{J}_3 \subset \mathfrak{J}_2 \subset \mathfrak{J}_1$.

If $G \varepsilon \mathfrak{J}_2$ and $q \varepsilon G$, there is a neighborhood $V_{p_o}(g_1,\ldots,g_k;\ \epsilon_1,\ldots,\epsilon_k)$ containing q and contained in G. Let $\{p_n\}$ be a sequence in D with $p_n \rightarrow p_o$ in the $\mathfrak{J}_1$ topology. By Theorem B.1(b), $\int g_i dp_n \rightarrow \int g_i dp_o$, i=1,...,k, and so there exists an index n_o such that

$$|\int g_i dp_{n_o} - \int g_i dp_o| < 1/2[\epsilon_i - |\int g_i dp_o - \int g_i dq|],\ i=1,\ldots,k,$$

Choose ϵ_i' rational such that

$$|\int g_i dp_{n_o} - \int g_i dp_o| + |\int g_i dp_o - \int g_i dq| < \epsilon_i' < 1/2[\epsilon_i + |\int g_i dp_o - \int g_i dq|],$$

i=1,...,k. Then

$$q \varepsilon V_{p_{n_o}}(g_1,\ldots,g_k;\ \epsilon_1',\ldots,\epsilon_k') \subset V_{p_o}(g_1,\ldots,g_k;\ \epsilon_1,\ldots,\epsilon_k),$$

and so $G \varepsilon \mathfrak{J}_3$. Therefore $\mathfrak{J}_2 = \mathfrak{J}_3$.

Let G be in $\mathfrak{J}_1$. Then X-G is closed relative to $\mathfrak{J}_1$. Let p be a limit point of X-G relative to $\mathfrak{J}_2$, i.e. there exists a net $\{p_\alpha\}$ in X-G such that $p_\alpha \longrightarrow p$ relative to $\mathfrak{J}_2$. Then for every k,

$$\int g_k dp_\alpha \rightarrow \int g_k dp$$

This net contains a sequence $\{p_n\}$ for which

$$\int g_k dp_n \rightarrow \int g_k dp$$

for every k, and by choice of the g_k's, $p_n \longrightarrow p$ relative to $\mathfrak{J}_1$. Therefore p is a limit point of X-G relative to $\mathfrak{J}_1$ and so is in X-G. This proves X-G is closed relative to $\mathfrak{J}_2$ and $G \varepsilon \mathfrak{J}_2$. It follows that $\mathfrak{J}_1 = \mathfrak{J}_2$. QED

<u>Theorem B.3</u> Let X be a Borel space and $\mathcal{J}$ a class of subsets of X which generates the σ-algebra $\mathcal{B}_X$. Then $\mathcal{B}_{P(X)}$ is the smallest σ-algebra on P(X) with respect to which the mappings

$$\theta_B\colon p \longrightarrow p(B),\ B \in \mathcal{J},$$

are measurable.

Proof:

Let Σ^* be the smallest σ-algebra for which the mappings θ_B, $B \in \mathcal{J}$, are measurable.

Let $\mathcal{E}$ be the class of Borel sets E in X for which θ_E is $\mathcal{B}_{P(X)}$ measurable. We show $\mathcal{E}$ is a Dynkin system:

(a) The space X is in $\mathcal{E}$, since p(X)=1 for every p.

(b) If $A, B \in \mathcal{E}$ and $B \subset A$, then for Q the set of rational numbers and c real,

$$\{p\colon p(A-B) < c\} = \bigcup_{r \in Q} \{p\colon p(A) < c+r,\ p(B) > r\}$$

is $\mathcal{B}_{P(X)}$ measurable. Therefore $A-B \in \mathcal{E}$.

(c) If $A_1, A_2, \ldots$ are in $\mathcal{E}$ and $A_n \uparrow A$, then

$$\{p\colon p(A) \leq c\} = \bigcap_{n=1}^{\infty} \{p\colon p(A_n) \leq c\}$$

is $\mathcal{B}_{P(X)}$ measurable. Therefore $A \in \mathcal{E}$.

This shows $\mathcal{E}$ is a Dynkin system.

Now let F be a closed subset of X and let d be a metric on X consistent with its topology. If F=X, then $F \in \mathcal{E}$. If $F \neq X$, define

$$F_n = \{x\colon d(x,F) \geq 1/n\},\ n=1,2,\ldots,$$

where $d(x,F) = \inf_{y \in F} d(x,y)$. For n sufficiently large, F_n is nonempty.

When F_n is nonempty, the nonnegative function

$$f_n(x) = d(x,F_n)/[d(x,F_n) + d(x,F)]$$

is in $C(X)$, is identically one on F and is identically zero on F_n. For n sufficiently large, choose $p \in P(X)$ which assigns mass one to some element in F_n. For $c>0$,

$$\{q: \int f_n dq < c\} = V_p(f_n; c)$$

is open in $P(X)$, so the mappings $q \to \int f_n dq$ are $\mathcal{B}_{P(X)}$-measurable. The mapping θ_F is a monotone limit of these mappings, so $F \in \mathcal{E}$.

The class of closed subsets of X is closed under finite intersections and is contained in $\mathcal{E}$. By the Dynkin system theorem [1, Theorem 4.1.2], $\mathcal{E} = \mathcal{B}_X$. This proves $\Sigma^* \subset \mathcal{B}_{P(X)}$.

To prove the reverse containment, it suffices to prove that $\mathcal{U}_3 \subset \Sigma^*$, where $\mathcal{U}_3$ is the countable basis for the topology on $P(X)$ defined in Theorem B.2. To show this, it suffices to prove that the sets of the form $V_p(g;\epsilon)$ are in Σ^*. But

$$V_p(g;\epsilon) = \{q: |\int g\, dq - \int g\, dp| < \epsilon\},$$

and since $q \to \int g\, dq$ is Σ^* measurable [9, (2.2)], $V_p(g;\epsilon)$ is also. QED

<u>Corollary B.3.1</u> Let X be a Borel space and $(\Omega, \mathcal{A})$ a measure space. Let Φ be a measurable map from $(\Omega, \mathcal{A})$ to $P(X)$ and let f be a function from ΩX to the extended real numbers, measurable with respect to the product σ-algebra $\mathcal{A}\mathcal{B}_X$. Assume f is bounded or f is nonnegative. Then the mapping

$$\omega \to \int_X f(\omega,x)\Phi(\omega)(dx)$$

is measurable from $(\Omega, \mathcal{A})$ to the extended real numbers.

Proof:

If f is bounded, this follows from Theorem B.3 and [9], (2.2). If f is nonnegative, let $\{f_n\}$ be a sequence of bounded measurable functions converging up to f. The mappings $\omega \to \int_X f_n(\omega, x)\, \Phi(\omega)(dx)$ are measurable, so their limit is also. QED

Corollary B.3.2 Let X be a Borel space and f a measurable function from X to the extended real numbers. Assume f is bounded or f is nonnegative. Then the mapping

$$p \to \int f \, dp$$

is measurable from P(X) to the extended real numbers.

Proof:

Let $\Omega = P(X)$, $\mathcal{A} = \mathcal{B}_{P(X)}$, and $\Phi(p) = p$ in Corollary B.3.1. QED

Corollary B.3.3 Let X be a Borel space, $\mathcal{J}$ a class of subsets of X which generates the σ-algebra $\mathcal{B}_X$, and $(\Omega, \mathcal{A})$ a measure space. A mapping Φ from $(\Omega, \mathcal{A})$ to P(X) is measurable if and only if the mappings

$$\omega \to \Phi(\omega)(B), \; B \in \mathcal{J},$$

are measurable.

Proof:

The mapping $\omega \to \Phi(\omega)(B)$ is $\theta_B \circ \Phi$. If Φ is measurable, the composition is also. If the composition is measurable, then $\Phi^{-1}\{\theta_B^{-1}[0,c)\}$ is in $\mathcal{A}$ for each $c>0$ and $B \in \mathcal{J}$. Since the sets of the form $\theta_B^{-1}[0,c)$, $B \in \mathcal{J}$, generate $\mathcal{B}_{P(X)}$, Φ is measurable. QED

<u>Lemma B.1</u> Let D be a dense subset of a metric space X and g a uniformly continuous function from D into a complete metric space Z. Then g has a unique extension to a continuous function on X.

Proof:

For $x \notin D$, let $\{x_k\} \subset D$ be such that $x_k \to x$. By the uniform continuity of g, $\{g(x_k)\}$ is Cauchy in Z, so g(x) can be defined as $\lim_k g(x_k)$. Use the uniform continuity of g again to show this extension is well-defined. It is clearly continuous and unique. QED

<u>Definition B.1</u> Let X and Y be Borel spaces and $\varphi: X \to Y$ be a Borel measurable one-to-one function such that $\varphi(X)$ is in $\mathcal{B}_Y$ and φ^{-1} is Borel measurable. Then φ is said to be a <u>Borel isomorphism</u>, and X and $\varphi(X)$ are <u>Borel isomorphic</u>.

As a result of the following theorem [26, Chapter I, Corollary 3.3], we need only check that φ in Definition B.1 is Borel measurable and one-to-one to conclude that it is a Borel isomorphism.

<u>Theorem B.4</u> (Kuratowski Theorem) If X and Y are Borel spaces and $\varphi: X \to Y$ is Borel measurable and one-to-one, then $\varphi(X)$ is Borel in Y and φ^{-1} is Borel measurable.

<u>Theorem B.5</u> Let X and Y be Borel spaces and $\theta: X \to Y$ a homeomorphism.[1] Then P(X) is homeomorphic to a Borel subset of P(Y).

Proof:

Define $\varphi: P(X) \to P(Y)$ by

[1] A homeomorphism is a topology-preserving one-to-one mapping of one topological space into another. We do not require it to be onto.

$$\Phi(p)(B) = p(\theta^{-1}(B)),\ B\epsilon\mathcal{B}_Y.$$

The mapping Φ is a one-to-one mapping of $P(X)$ onto $\{q\epsilon P(Y):\ q(\theta(X))=1\}$ and this is in $\mathcal{B}_{P(X)}$ by Theorems B.3 and B.4.

If $\{p_n\}$ is a sequence in $P(X)$ converging to $p\epsilon P(X)$, and f is in $C(Y)$, then $f\circ\theta$ is in $C(X)$ and

$$\int_Y f\ d\Phi(p_n) = \int_X (f\circ\theta)dp_n \to \int_X (f\circ\theta)dp = \int_Y f\ d\Phi(p),$$

so Φ is continuous.

Let d be a metric on Y consistent with its topology and define a metric d' on X by

$$d'(x_1,x_2) = d(\theta(x_1),\theta(x_2)),\ x_1,x_2\epsilon X.$$

Since θ is a homeomorphism, d' is consistent with the topology on X. Let g' be in $U_{d'}(X)$. Then $g=g'\circ\theta^{-1}$ is in $U_d(\theta(X))$. By Lemma B.1 and the Tietze extension theorem [11], g can be extended to $g^*\epsilon\, C(Y)$. If $\{p_n\}$ is a sequence in $P(X)$ and for some $p\epsilon P(X)$, $\Phi(p_n)\to\Phi(p)$ in $P(Y)$, then

$$\int_X g'dp = \int_{\theta(X)} g\ d\Phi(p_n) = \int_Y g^* d\Phi(p_n) \to \int_Y g^* d\Phi(p) = \int_X g'dp,$$

so Φ^{-1} is continuous. QED

<u>Theorem B.6</u> If Y is a complete separable space, then $P(Y)$ is a complete separable space.

See [26], Chapter II, Theorems 6.2 and 6.5 for a proof.

<u>Corollary B.6.1</u> If X is a Borel space, then $P(X)$ is also.

Proof:

By Theorems B.5 and B.6, $P(X)$ is homeomorphic to a Borel subset of a complete separable space. QED

Theorem B.7 Let X be a Borel space and $\bar{X} = \{p_x: x \in X\}$, where p_x is the probability assigning mass one to x. Then X is homeomorphic to $\bar{X}$ and $\bar{X}$ is a Borel subset of $P(X)$.

Proof:

Define $\theta(x) = p_x$. Let $x_k \to x$ and G be open in X. If $x \in G$, then for sufficiently large k, $x_k \in G$ and

$$\liminf_k p_{x_k}(G) = 1 = p_x(G).$$

If $x \notin G$, then

$$\liminf_k p_{x_k}(G) \geq 0 = p_x(G).$$

Therefore $p_{x_k} \to p_x$ and θ is continuous by Theorem B.1(e).

Now let p_{x_k} be a sequence converging to p_x in $\bar{X}$. Let G be an open set containing x. By Theorem B.1(e),

$$\liminf_k p_{x_k}(G) \geq p_x(G) = 1,$$

and so $x_k \in G$ for all k sufficiently large. This implies $x_k \to x$ and θ^{-1} is continuous.

The set $\bar{X}$ is Borel by Theorem B.4. QED

We separate out a part of the proof of Theorem B.7 as a corollary.

Corollary B.7.1 Let X be a Borel space. The mapping $x \to p_x$ is continuous from X to $P(X)$.

We now prove a lemma (given here as Lemma B.4) found in [6]. The proof given there contains an error which is corrected below.

If X is a compact metric space, 2^X is the set of (possibly empty) closed subsets of X. If d is the metric in X and K_1 and K_2 are nonempty closed subsets of X, we define the <u>Hausdorff metric</u> [15, Section 28] by

$$d(K_1,K_2) = \max\{\sup_{x\in K_1} d(x,K_2),\ \sup_{x\in K_2} d(x,K_1)\},$$

where

$$d(x,K) = \inf_{y\in K} d(x,y).$$

If $K_1=\emptyset$ and K_2 is a nonempty closed subset of X, define

$$d(K_1,K_2) = d(K_2,K_1) = \text{diameter}(X),$$

where $\text{diameter}(X)=\sup_{x,y\in X} d(x,y)$ is finite because X is compact. Thus the empty set is an isolated element of 2^X. This metric gives rise to the <u>exponential topology</u> on 2^X ([19], Section 17; [20], Section 43) in which 2^X is a compact metric space.

We denote by N the countable cross product of the set of positive integers. Let the set of positive integers have the discrete topology and N the product topology. The space N has a metrization consistent with this topology. If z is an element of N, z_i will be its i-th component. A sequence in N will be indicated by z(1), z(2), ...

We define functions from N to the power set of N by

$$L(z) = \{w\in N:\ w_i \le z_i,\ i=1,2,\ldots\}$$
$$N_k(z) = \{w\in N:\ w_i = z_i,\ i=1,2,\ldots,k\}.$$

For each $z \in N$, $L(z)$ is a compact set by Tychonoff's Theorem.

<u>Lemma</u> <u>B.2</u> Let f be a continuous function from N to a metric space X. For $z \in N$,

$$\sup_{y \in L(z)} \text{diameter}(f[N_k(y)]) \to 0 \text{ as } k \to \infty.$$

Proof:

Suppose the contrary. Since $N_k(y) \supset N_{k+1}(y)$, the expression above is monotone decreasing in k. For some $\epsilon > 0$ and every k there is a $y_k \in L(z)$ such that

$$\text{diameter}(f[N_k(y)]) \geq \epsilon.$$

Let $\Pi_k = \{y \in L(z): \text{ diameter}(f[N_k(y)]) \geq \epsilon\}$. Then $\Pi_1 \supset \Pi_2 \supset \ldots$ and $\Pi_k \neq \emptyset$ for each k.

Since $N_k(y)$ depends only on the first k components of y, the set

$$\{y \in N: \text{ diameter}(f[N_k(y)]) < \epsilon\}$$

is open in N, and so Π_k is closed, therefore compact. It follows that there exists $y^* \in \bigcap_{k=1}^{\infty} \Pi_k$. For each k, there exists $z(k) \in N_k(y^*)$ with

$$d(f(z(k)), f(y^*)) \geq \epsilon/3.$$

But $z(k) \to y^*$ and f is continuous. This leads to a contradiction. QED

<u>Lemma</u> <u>B.3</u> Let X be a complete separable metric space and f a continuous function from N to X. For $p \in P(X)$,

$$p[f(N)] = \sup_{z \in N} p(f[L(z)]).$$

Proof:

Since for each z, L(z) is compact, f(L(z)) is compact and thus in Φ_X. By Corollary A.5.1, the p measure of f(N) can be defined.

For each z,

$$f(N) \supset f[L(z)],$$

so

$$p[f(N)] \geq \sup_{z \in N} p(f[L(z)]).$$

Define

$$N(n_1,\ldots,n_k) = \{w \in N: w_1 \leq n_1,\ldots,w_k \leq n_k\},$$
$$A(n_1,\ldots,n_k) = f[N(n_1,\ldots,n_k)].$$

Since $N(n_1,\ldots,n_k)$ is open in N, it is Borel and $A(n_1,\ldots,n_k)$ is analytic. The increasing sequence $\{N(n_1,\ldots,n_k,1),\ N(n_1,\ldots,n_k,2),\ \ldots\}$ converges to $N(n_1,\ldots,n_{k-1})$, so the sequence $\{A(n_1,\ldots,n_k,1),\ A(n_1,\ldots,n_k,2),\ \ldots\}$ is also increasing and converges to $A(n_1,\ldots,n_{k-1})$. The sequence $\{A(1),\ A(2),\ \ldots\}$ increases to f(N).

Given $\epsilon>0$, construct a sequence $z = (z_1,\ z_2,\ \ldots)$ of positive integers for which

$$p[f(N)] \leq p[A(z_1)] + \epsilon/2,$$
$$p[A(z_1,\ldots,z_{k-1})] \leq p[A(z_1,\ldots,z_k)] + \epsilon/2^k.$$

Then $f(N) \supset A(z_1) \supset A(z_1,z_2) \supset \ldots$, and defining

$$A(z) = \bigcap_{k=1}^{\infty} A(z_1,\ldots,z_k),$$

we have

$$p[f(N)] \leq p[A(z)] + \epsilon.$$

To conclude the proof, we show $A(z) = f[L(z)]$.

Clearly $A(z) \supset f[L(z)]$, so we show the reverse containment. Suppose $x \epsilon A(z)$. Then for each k, there exists $y(k) \epsilon N(z_1, \ldots, z_k)$ such that $x = f(y(k))$. The sequence of positive integers $\{y(1)_1, y(2)_1, \ldots\}$ takes some value w_1 between 1 and z_1 infinitely often. Let I_1 be an infinite index set such that $y(j)_1 = w_1$ for $j \epsilon I_1$. The sequence $\{y(j)_2: j \epsilon I_1\}$ takes some value w_2 between 1 and z_2 infinitely often. Let $I_2 \subset I_1$ be an infinite index set such that $y(j)_2 = w_2$ for $j \epsilon I_2$. Continuing in this manner, construct a sequence of infinite index sets $I_1 \supset I_2 \supset I_3 \supset \ldots$ and $w \epsilon L(z)$ such that $y(j)_k = w_k$ for $j \epsilon I_k$, $k=1,2,\ldots$. Choose $m_1 \epsilon I_1$, $m_2 \epsilon I_2$, ... such that $m_1 < m_2 < \ldots$. The sequence $y(m_1)$, $y(m_2)$, ... converges to w, and since f is continuous, $x = f(w)$. Therefore $x \epsilon f[L(z)]$. QED

<u>Lemma B.4</u> Let X be a compact metric space and A an analytic subset of X (see Appendix A). Let c be a real number. Then

$$\{p \epsilon P(X): p(A) > c\}$$

is analytic.

Proof:

Since A is analytic, there is a continuous function $f: N \rightarrow X$ such that $f(N) = A$. We show that the function

$$z \rightarrow f[L(z)]$$

is continuous from N to 2^X.

Note that since $L(z)$ is compact and f is continuous, $f[L(z)]$ is in 2^X. Let $z(n) \to z$ in N. Choose $\epsilon > 0$. By Lemma B.2, there exists k such that

$$\sup_{y \in L(z)} \text{diameter}(f[N_k(y)]) < \epsilon.$$

There exists N_k such that $n \geq N_k$ implies $z(n)_i = z_i$, $i=1,2,\ldots k$. Suppose $y \in L(z)$. Let $y(n)_i = \min\{z(n)_i, y_i\}$. For $n \geq N_k$, $y(n) \in N_k(y)$ and

$$d(f(y), f(y(n))) < \epsilon,$$

where d is the metric on X. Therefore

$$\sup_{x \in f[L(z)]} d(x, f[L(z(n))]) \to 0 \tag{B.1}$$

as $n \to \infty$.

For $w(n) \in L(z(n))$, define $w \in L(z)$ by

$$w_i = \min\{w(n)_i, z_i\}.$$

Then as before for $n \geq N_k$, we have $w(n) \in N_k(w)$ and

$$d(f(w(n)), f(w)) < \epsilon.$$

Therefore

$$\sup_{x_n \in f[L(z(n))]} d(x_n, f[L(z)]) \to 0 \tag{B.2}$$

as $n \to \infty$. Relations (B.1) and (B.2) imply $f[L(z(n))] \to f[L(z)]$ in 2^X and this establishes continuity.

By [9, (3.8)], the mapping

$$(p, K) \to p(K)$$

is upper semicontinuous from $P(X)2^X$ to $[0,1]$. Therefore the composition

$$(p,z) \longrightarrow p(f[L(z)])$$

is upper semicontinuous, i.e.

$$B(c) = \{(p,z)\colon p(f[L(z)]) \geq c\}$$

is closed in $P(X)N$ for each real c. Then

$$B = \{(p,z)\colon p(f[L(z)]) > c\} = \bigcup_{k=1}^{\infty} B(c + 1/k)$$

is Borel in $P(X)N$, and by Lemma B.3

$$\{p \in P(X)\colon p(A) > c\}$$

is the projection of B on the $P(X)$-axis. This projection is analytic by Theorem A.3. QED

Note that by Theorem A.2, the following statements are equivalent:

(a) $\{p\colon\ p(A) > c\}$ is analytic for each real c;

(b) $\{p\colon\ p(A) \geq c\}$ is analytic for each real c.

We will use (a) and (b) interchangeably.

We now extend Lemma B.4 to the noncompact case.

<u>Theorem B.8</u> Let X be a Borel set and A an analytic subset of X. Let c be a real number. Then $\{p \in P(X)\colon\ p(A) > c\}$ and $\{p \in P(X)\colon\ p(A) \geq c\}$ are analytic.

Proof:

By Urysohn's Theorem [11, Chapter IX, Corollary 9.2], there is a homeomorphism θ which embeds X in a compact metric space Y. Let $\Phi\colon P(X) \longrightarrow P(Y)$ be the homeomorphism defined in Theorem B.5. By Theorem A.3, $\theta(A)$ is

analytic. By Lemma B.4, $\{q \varepsilon P(Y): q(\theta(A)) > c\}$ is analytic. Again by Theorem A.3,

$$\{p \varepsilon P(X): p(A) > c\} = \varphi^{-1}(\{q \varepsilon P(Y): q(\theta(A)) > c\})$$

is analytic. The remainder of the theorem follows from Theorem A.2. QED

APPENDIX C

STOCHASTIC KERNELS

Definition C.1 Let $(X,\mathcal{A})$ and $(Y,\mathcal{B})$ be measure spaces. Then $q(dy|x)$ is an $\mathcal{A}$-measurable stochastic kernel on Y given X if

(a) $q(\cdot|x)$ is a probability measure on $(Y,\mathcal{B})$ for each x;

(b) $q(E|\cdot)$ is $\mathcal{A}$-measurable for each $E \in \mathcal{B}$.

When there is no possibility of confusion, q will be called simply a stochastic kernel. If Y is a product space $Y = WZ$, we will write $q(dw \cdot Z|x)$ to denote the stochastic kernel which, for each x, is the marginal of $q(dy|x)$ on W. Similarly the marginal on Z will be represented by $q(W \cdot dz|x)$. If p is a measure on $(Y,\mathcal{B})$, we will write $p(dw \cdot Z)$ and $p(W \cdot dz)$ to indicate the marginals of p on W and Z respectively.

The two major existence theorems concerning stochastic kernels follow. The first is well-known and states that in a product of Borel spaces, any measure can be decomposed into its marginal and a stochastic kernel. The second is a generalization of the first to include a measurable dependence on a parameter.

Theorem C.1 Let X and Y be Borel spaces and p an element of $P(XY)$. Then there exists a stochastic kernel $q(dy|x)$ such that

$$P(\underline{X}\underline{Y}) = \int_{\underline{X}} q(\underline{Y}|x)\, p(dx \cdot Y),$$

for every $\underline{X} \in \mathcal{B}_X$, $\underline{Y} \in \mathcal{B}_Y$.

Proof:

This is an easy consequence of [1], Theorems 6.6.5 and 6.6.6. As a result of these theorems, given $p \in P(XY)$ and a σ-algebra $\mathcal{D} \subset \mathcal{B}_{XY}$, there exists

a regular conditional probability $Q((x,y),B)$, i.e. a function on $XY\mathcal{B}_{XY}$ which is $\mathcal{J}$-measurable in (x,y) for fixed $B\in\mathcal{B}_{XY}$ and a probability measure on $(XY,\mathcal{B}_{XY})$ for fixed (x,y). If $\mathcal{J}$ is the σ-algebra of sets of the form $\underline{X}Y$, $\underline{X}\in\mathcal{B}_X$, then $Q((x,y,),B)$ is independent of y. Define

$$q(\underline{Y}|x) = Q((x,y),X\underline{Y}),$$

where y is arbitrary. Then by the defining property of conditional probabilities

$$P(\underline{XY}) = \int_{\underline{X}Y} Q((x,y),X\underline{Y})\ p(d(x,y)) = \int_{\underline{X}} q(\underline{Y}|x)\ p(dx\cdot Y)$$

for $\underline{X}\in\mathcal{B}_X$, $\underline{Y}\in\mathcal{B}_Y$. QED

<u>Theorem</u> <u>C.2</u> Let X and Y be Borel spaces and $(\Omega,\mathcal{A})$ a measure space. Let $p(d(x,y)|\omega)$ be a stochastic kernel on XY given Ω. Then there exists a $\mathcal{B}_X\mathcal{A}$-measurable stochastic kernel $q(dy|x,\omega)$ such that

$$p(\underline{XY}|\omega) = \int_{\underline{X}} q(\underline{Y}|x,\omega)\ p(dx\cdot Y|\omega)$$

for every $\underline{X}\in\mathcal{B}_X$, $\underline{Y}\in\mathcal{B}_Y$.

Proof:

The theorem follows from [36], Lemma A.1.9, when X and Y are the real line. This can be extended to allow X and Y to be Borel subsets of the unit interval. By Theorem B.4 and [1], Theorem 6.6, every Borel space is Borel isomorphic to a Borel subset of the unit interval. QED

<u>Theorem</u> <u>C.3</u> Let X be a Borel space and $(\Omega,\mathcal{A})$ a measure space. A function $q(\underline{X}|\omega)$ on $\mathcal{B}_X\Omega$ is an $\mathcal{A}$-measurable stochastic kernel if and only if for each ω, $q(\cdot|\omega)\in P(X)$ and the mapping

$$\omega \longrightarrow q(\cdot|\omega)$$

is $\bigstar$-measurable.

Proof:

This is a direct consequence of Corollary B.3.3. QED

REFERENCES

[1] R. Ash, Real Analysis and Probability, Academic Press, New York, 1972.

[2] R. Bellman, Dynamic Programming, Princeton University Press, Princeton, New Jersy, 1957.

[3] D. P. Bertsekas, Dynamic Programming and Stochastic Control, Academic Press, New York, 1976.

[4] D. P. Bertsekas, Monotone mappings with applications in dynamic programming, SIAM J. Control and Optimization 15 (1977), to appear.

[5] D. Blackwell, Discounted dynamic programming, Ann. Math. Stat. 36 (1965), 226-235.

[6] D. Blackwell, D. Freedman, and M. Orkin, The optimal reward operator in dynamic programming, Ann. Prob. 2 (1974), 926-941.

[7] L. D. Brown and R. Purves, Measurable selections of extrema, Ann. Stat. 1 (1973), 902-912.

[8] E. V. Denardo, Contraction mappings in the theory underlying dynamic programming, SIAM Review 9 (1967), 165-177.

[9] L. Dubins and D. Freedman, Measurable sets of measures, Pacific J. Math. 14 (1964), 1211-1222.

[10] L. Dubins and L. Savage, Inequalities for Stochastic Processes (How to Gamble if you Must), Dover, New York, 1976.

[11] J. Dugundji, Topology, Allyn and Bacon, Boston, 1966.

[12] E. B. Dynkin and A. A. Juskevic, Controlled Markov Processes and their Applications, Moscow, 1975. English translation to be published by Springer-Verlag.

[13] D. Freedman, The optimal reward operator in special classes of dynamic programming problems, Ann. Prob. 2 (1974), 942-949.

[14] E. B. Frid, On a problem of D. Blackwell from the theory of dynamic

programming, Theory Prob. and Appl. (1970), 719-722.

[15] F. Hausdorff, Set Theory, Chelsea, New York, 1957.

[16] K. Hinderer, Foundations of Nonstationary Dynamic Programming with Discrete Time Parameter, Springer-Verlag, Berlin, 1970.

[17] R. A. Howard, Dynamic Programming and Markov Processes, Wiley, New York, 1960.

[18] K. Kuratowksi and C. Ryll-Nardzewski, A general theorem on selectors, Bull. Polish Acad. Sci. 13 (1965), 397-411.

[19] K. Kuratowski, Topology I, Academic Press, New York, 1966.

[20] K. Kuratowski, Topology II, Academic Press, New York, 1968.

[21] A. Maitra, Discounted dynamic programming on compact metric spaces, Sankhya 30A (1968), 211-216.

[22] J. von Neumann, On rings of operators. Reduction theory, Ann. of Math. (1949), 401-485.

[23] P. Olsen, When is a multistage stochastic programming problem well-defined? SIAM J. Control and Optimization 14 (1976), 518-527.

[24] P. Olsen, Multistage stochastic programming with recourse as mathematical programming in an L_p space, SIAM J. Control and Optimization 14 (1976), 528-537.

[25] P. Olsen, Multistage stochastic programming with recourse: The equivalent deterministic problem, SIAM J. Control and Optimization 14 (1976), 495-517.

[26] K. Parthasarathy, Probability Measures on Metric Spaces, Academic Press, New York, 1967.

[27] R. T. Rockafellar, Integral functionals, normal integrands and measurable selections. Published in Nonlinear Operators and the Calculus of Variations, Lucien Waelbroeck, ed., Springer-Verlag, 1976.

[28] R. T. Rockafellar and R. Wets, Stochastic convex programming: Relatively complete recourse and induced feasibility, SIAM J. Control and Optimization 14 (1976), 574-589.

[29] R. T. Rockafellar and R. Wets, Stochastic convex programming: Basic duality, Pacific J. Math., to appear.

[30] S. Saks, Theory of the Integral, G. E. Stechert and Co., New York, 1937.

[31] M. Schael, On continuous dynamic programming with discrete time parameter, Z. Wahrscheinlichkeitstheorie verw. Gebiete 21 (1972), 279-288.

[32] M. Schael, Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal, Z. Wahrscheinlichkeitstheorie verw. Gebiete 32, (1975), 179-196.

[33] M. Schael, On dynamic programming: Compactness of the space of policies, Stoch. Processes and their Appl. 3, (1975), 345-364.

[34] G. F. Simmons, Introduction to Topology and Modern Analysis, McGraw-Hill, New York, 1963.

[35] R. E. Strauch, Negative dynamic programming, Ann. Math. Stat. 37 (1966), 871-890.

[36] C. Striebel, Optimal Control of Discrete Time Stochastic Systems, Springer-Verlag, New York, 1975.

[37] A. Wald, Statistical Decision Functions, Wiley, New York, 1950.

[38] H. S. Witsenhausen, A standard form for sequential stochastic control, Math. Systems Theory 7, Springer-Verlag, New York, 1973.

VITA

Steven E. Shreve was born in Richwood, West Virginia, in 1950. He did his undergraduate work at West Virginia University, where he majored in German and minored in mathematics, graduating Magna Cum Laude in 1972. He was elected a member of Pi Mu Epsilon National Mathematics Honorary and Phi Beta Kappa while there. He then studied mathematics for a year at Georg-August-Universitaet, Goettingen, West Germany, as a Graduate Fellow of the Rotary Foundation. In 1973 he began graduate work in the Department of Mathematics of the University of Illinois and completed an M.S. degree in Electrical Engineering and a Ph.D. degree in Mathematics in January, 1977. Shreve has experience as a tutor employed by the Department of Mathematics of West Virginia University and as a teaching assistant for the Department of Mathematics of the University of Illinois. He is presently a Research Associate at the Coordinated Science Laboratory of the University of Illinois. With Dimitri Bertsekas, Shreve has co-authored the papers:

Equivalent stochastic and deterministic optimal control problems, Proc. IEEE Conf. on Decision and Control, Clearwater Beach, Florida (1976), 705-709.

A new theoretical framework for finite horizon stochastic control, Proc. 14th Annual Allerton Conf. on Circuit and System Theory, Allerton Park, Illinois (1976).